CHAPTER 4.2: APPLICATION OF URBANSIM IN PARIS (ILE-DE-FRANCE) CASE STUDY

André de Palma, Nathalie Picard, Kiarash Motamedi

ECOLE POLYTECHNIQUE

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

CHAPTER 4.2: APPLICATION OF URBANSIM IN PARIS (ILE-DE-FRANCE) CASE STUDY

André de PALMA
Nathalie PICARD
Kiarash MOTAMEDI

December 2014

Cahier n° 2014-33

DEPARTEMENT D'ECONOMIE
Route de Saclay
91128 PALAISEAU CEDEX
(33) 1 69333033
http://www.economie.polytechnique.edu/
mailto:chantal.poujouly@polytechnique.edu

Chapter 4.2: Application of UrbanSim in Paris (Ile-de-France) Case study

André de Palma*[1], Nathalie Picard†[2] and Kiarash Motamedi‡[3]

[1]CES, Ecole Normale Supérieure de Cachan, Cachan, France and CECO, École Polytechnique, Palaiseau, France
[2]THEMA, University of Cergy-Pontoise, Cergy, France and CECO, École Polytechnique, Palaiseau, France
[3]THEMA, University of Cergy-Pontoise, Cergy, France

December 1, 2014

Abstract

This chapter presents the modeling and simulation works applying UrbanSimE in Paris case study. This is the most important application of UrbanSim worldwide with respect to the size of the region and of the project evaluated, data availability and complexity. The rich available data allowed us to make different econometric analyses and to implement several theoretical insights previously presented in this handbook. The model has been successfully used in an applied study concerning the Grand Paris project, which can be considered as the most important investment plan in the transportation system of the region since the construction of Paris subway network at the beginning of the 20^{th} century.

1 Scope and Challenges

The integrated land use and transportation models have received increased attention in research and practice over the last decades. In particular, the researchers and the decision makers highlight the ability of these integrated models to examine the combined effects of land use and transport policies, to study the endogeneity of urban development and of travel patterns, as well as to analyze and quantify the effects of

*andre.depalma@ens-cachan.fr
†Corresponding author: nathalie.picard@u-cergy.fr
‡kiarash.motamedi@u-cergy.fr

transport capacity expansion. The usefulness of integrated models goes beyond the examination of induced effects of transport demand. These integrated tools allow examining issues such as efficiency and fairness in markets for local public goods, as well as for local public bads. In addition, integrated models are ideal tools to study spatial patterns and the distributional effects of environmental externalities related to congestion, such as emissions from vehicles, local pollution, noise and accidents. In this chapter, we use the integrated land use and transport model system UrbanSimE / METROPOLIS[1] to study the impact of a large transportation infrastructure project including the resulting inequalities in the spatial distribution of accessibility and environmental quality.

The main nine steps of the case study are described below. Each step has an interest in itself, but here we will mainly focus on the results not of each module, but of the overall integrated system.

1. Data collection and analysis of data;

2. definition of the appropriate geographical scales (*Commune, Département, IRIS, Ilots, Ilots MOS*[2] are different geographical units used in France). Such spatial issues were also studied from a theoretical point of view in Chapter 1.3 - see also Chapter 2.8, which deals with econometric issues. The terminology will be explained in details below in this chapter;

3. integration of a demographic model;

4. improvement and calibration of the real estate pricing model;

5. implementation of the firmography model;

6. improvement of the residential location choice model;

7. calibration of the transportation model -The Paris case study used METROPOLIS transportation model;

8. definition of scenarios and evaluation of such scenarios and

9. environmental and socio-economic policy analysis.

[1]The integration of two models is realized by an interface that manages the interaction between them and carries out the necessary data transformations. This interface is explained in chapter 3.3 of this handbook.

[2]*Ilot* refers to the basic census geographical unit the represents a block of buildings or land delimited by roads (streets). *IlotMOS* refers to the polygons with homogeneous land use cover represented in land use registry (*MOS*). *IRIS* refers to an aggregation of neighboring *Ilots* with about 2000 inhabitants or 1000 jobs.

2 Introduction

Among the three European case studies[3], Ile-de-France is the largest application of UrbanSimE (the European version of UrbanSim) with a total area of over 12,000 square kilometers and more than 5 million households and jobs. It has also benefited from the richest database, including exhaustive census data. This case study has also benefited from the results of several theoretical developments (see chapters 2.2, 2.3, 2.6 and SusutainCity working paper 3.5) . The results of this case study have been disseminated and used successfully in applied studies and socio-economic analyses, such as the Grand Paris project. We benefited from collaboration with the *Société du Grand Paris*, which helped us to define the scenarios to be considered in the context of the establishment of a network of rapid transit (Figure 1), that will be created in the coming decades. The French Prime Minister has recently presented the schedule for putting the different segments of the Grand Paris Express metro lines in operation. The first part of the project will be the southern part of the ring around the Paris. Other segments are completed consecutively to each other.

Several important questions should be answered. First, what should be the timing of the construction of different segments over the years to come? Second, what are the wide economic benefits? Third, how many jobs will be created (attracted) in the region in the coming years? Fourth, what will be the relocation decisions of different economic agents (residents and businesses)? This question is important because it will help the local authorities to plan to cope with the demand for housing (especially regarding collective dwellings) and for offices. And the fifth question: What will be the traffic demand, both in the existing and new public transport network and on roads?

This chapter has 10 sections. The 8 remaining sections are presented as follows:

Section 3 presents the study area. Section 4 presents data collection. This is the first key task, since the quality of calibration and the quality of the prediction mainly depend on the quality of the data. Data were collected to estimate the four modules (residential location, firm location, real estate and land use) and to calibrate the results produced by the integrated transport-land-use model. It covers demography, population (exhaustive census data), firms and jobs (exhaustive Regional Employment Surveys), yearly real-estate prices in the largest 300 *communes*, land-use and many other marginal data sources. Moreover, transportation data (supply and demand) were collected. All the data needed to run UrbanSimE were collected. All these data have been carefully checked. The analysis of data is discussed in Section 5. The structure of the model and the description of the different sub-models (residential location, firmography, real estate and land use) are discussed in Section 6. The transportation model, METROPOLIS is presented in Section 7. The integrated

[3]Paris (Ile-de-France, Zurich and Brussels constitute the 3 case studied of SustainCity project. The other cases are presented respectively in chapters 4.3 and 4.4. To the best of our knowledge, Ile-de-France is even the worldwide largest application of UrbanSim.

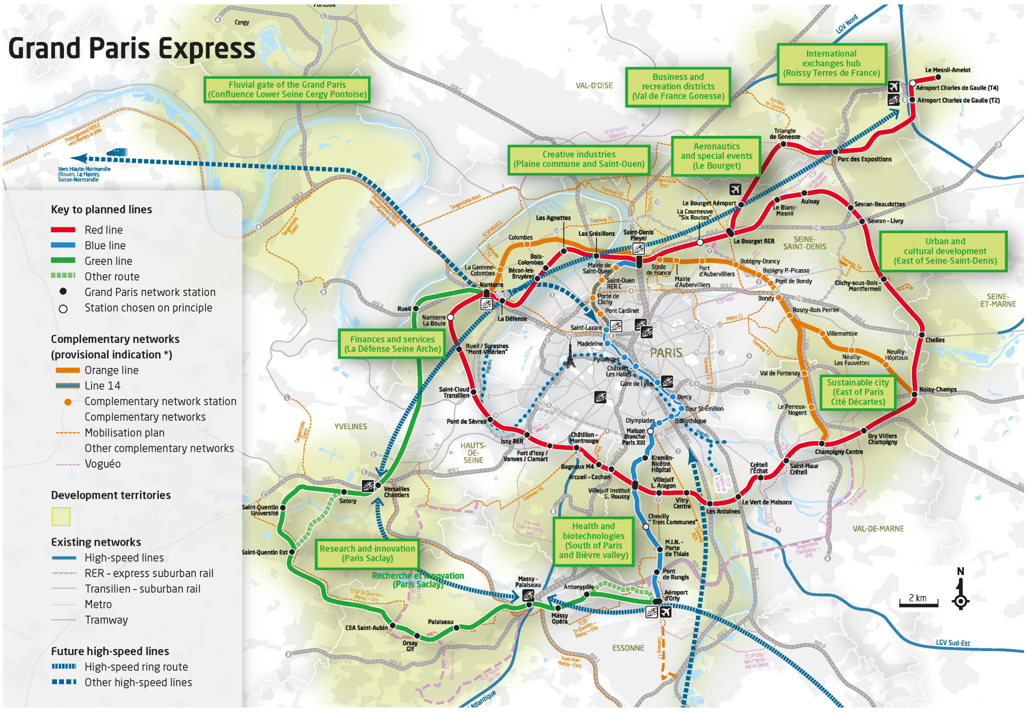


Figure 1: The Grand Paris Railway Network (Source: **?**)

land use and transportation model (detailed in Chapter 3.3) is summarized in section 8. The major part of this chapter focuses on the results (Section 9). To the best of our knowledge, this is the first wide application of UrbanSim used to address major issues faced in a large metropolitan area. The description of policy scenario of the Grand Paris project as well as the results are presented in Section 9. Section 10 concludes.

3 Study area

The Paris area, namely Ile-de-France Region, embraces Paris and its suburbs. It is Europe's most populated region. The city of Paris has about 2 million inhabitants, on a regional total of 11 million. The total number of jobs is 5.1 million. With a surface of 12,000 sq. km, Ile-de-France Region occupies 2.2% of the surface of France. It represents 19% of the population, 22% of the jobs and 29% of the GDP of the country. The GDP is at the level of EUR 552,100 million which is the fifth-largest in the world, just before London, and after the urban areas of Tokyo, New York, Los Angeles and Chicago. Ile-de-France is the Europe's biggest employment base and concentrates over 50% of French executives living in the area. The large fraction of senior professionals is due to the high density of company headquarters located in Paris Region. However, Paris Region economy is also diversified.

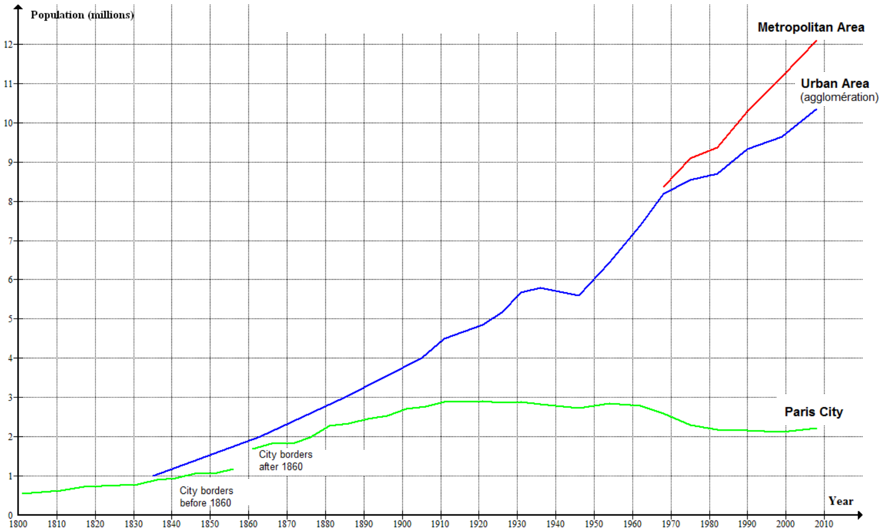


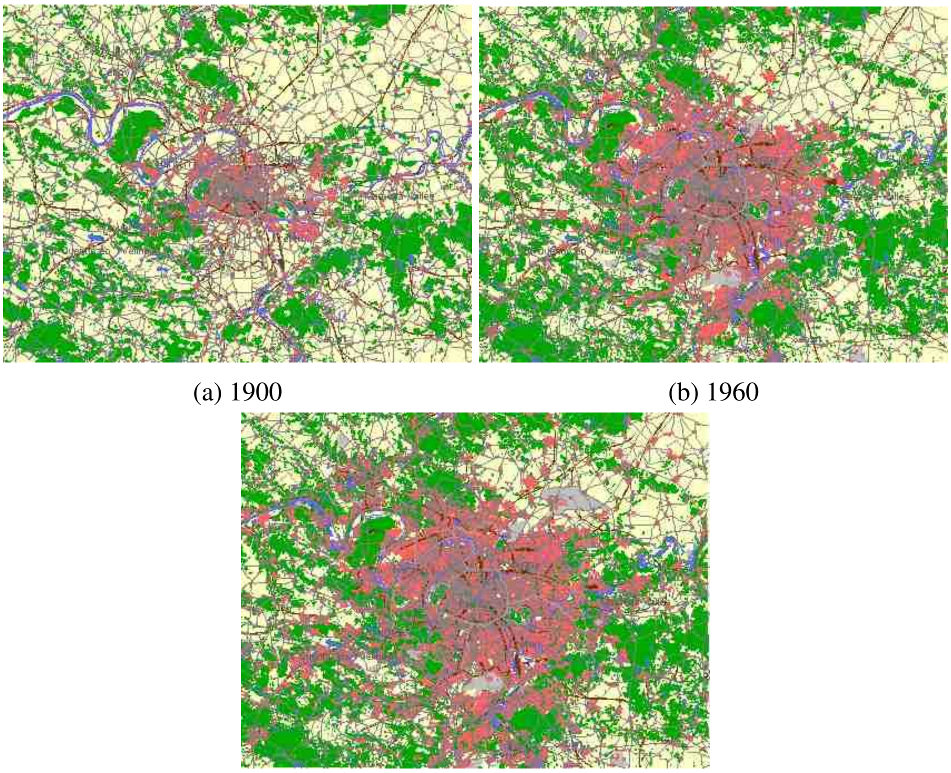
Figure 2: Population of Paris and its suburbs from 1801 to 2008 (Source: **?**)

There are 3 administrative institutions in Ile-de-France: 1 “*Région*”, 8 “*Départements*” (Counties) and 1300 “*Communes*” (Municipalities). In addition, we consider the 3 counties around Paris (*Hauts-de-Seine, Val-de-Marne* and *Seine-Saint-Denis*) as close suburb or “small ring” and the 5 counties far away from Paris (*Yvelines, Essonne, Val-d’Oise* and *Seine-et-Marne*) as far suburb or “large ring”. Very large differences in population density can be observed between Paris and the outer periphery.

The genesis of Paris begins with "l’Ile de la Cité" a small island at the center of Paris city built, 56 BC with surface of 2ha. The population of Paris city reached its maximum at 2.9 million habitants in 1920. The borders of the current Paris city have been fixed in then 1930’s with a surface of 10,54 kha. Its population is about 2.26 million inhabitants in 2012. Figure 2 represents the evolution of the population of Paris city and its area. The population of Paris has been decreasing during the 1960’s and 1970’s then remained steady up to 2000. Meanwhile, the population of region increased rapidly when it began to locate around the secondary poles and not directly in the influence zone of Paris. Figure 3 represents the evolution of the urbanization in the Paris area. The urbanized area was concentrated in Paris and some nearby poles in 1900, occupied its main space in 1960 and then expanded its boundaries up to 2000.

Based on the English experience of new towns or *Villes nouvelles*[4] during the

[4] *Ville Nouvelle* is a French administrative terminology corresponding to 4 cities that include Cergy-Pontoise, Marne-la-vallée, Sénart, Saint-Quentin-en-Yvelines and Evry. They are connected to CBD by fast public transportation (RER) lines. They are aimed to be the secondary centers in order to decentralize population and jobs. As a consequence, the Paris area can be seen as a polycentric city



(a) 1900

(b) 1960

(c) 2000

Figure 3: Evolution of urbanization in Paris area (Source: *IAU-IdF*)

1960s and 1980s, the French government created *Villes nouvelles* in the outer ring of the Paris suburbs in order to multi-polarize the economy of the city (Combes et al. 2011). However, *Villes nouvelles* experiment showed the limits of the policy of developing population poles, since economic activities still remain in a large measure concentrated in the center of the urban area of Paris Region, namely, Paris City and *Hauts-de-Seine*. As illustrated in Figure 4 the new towns have well absorbed the overflow of Paris population growth. Other parts of the region keep growing as before. The growth of new towns seems to slow down as a consequence of the end of specific policies set by government at the mid of 2000's.

The land use is composed of built-up areas (30%), green areas (20%) and rural areas (50%). The public transportation network is made of:

- a main radial railway network, especially the RER lines (high speed train service between Paris and the suburbs);
- a subway network that provides comprehensive and timely service in the city centre;

with one major center (historical Paris) and several sub-centers.

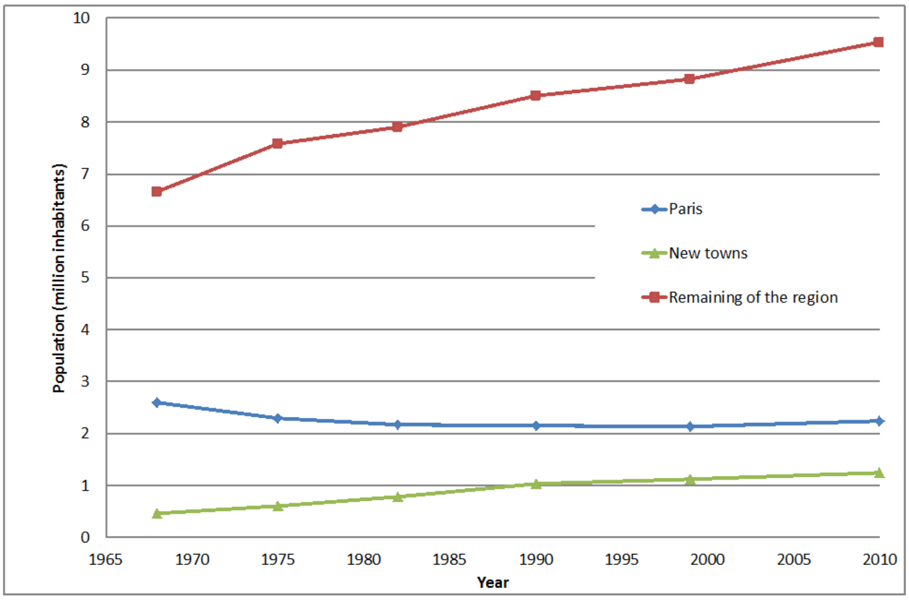


Figure 4: Population of Paris in comparison with new towns and rest of the region from 1968 to 2010 (Source: Census data, *INSEE*)

- a bus network to complement the rail services.

The road network is organized into a hierarchy that is densely interconnected and often congested. The express network of the region is composed of 590 km of motorways and 250 km of expressways, with a total of 4,500 lane-km. Road traffic flows reach the highest levels known all over the country. But thanks to an efficient road network, and despite the traditional traffic jams (at rush-hour), traffic conditions are on the whole remarkably good for a metropolis of this size, since the average duration of a car trip is 19 minutes, and no more than 25 minutes for commuting trips by car (DREIF 2005). The market shares for the home based work trips in 2001 were 50% Private cars, 36% Transit and 14% Bicycle/walk (and other modes). Over the last twenty years, the use of public transportation has decreased down to 6% in the region. This can be explained by structural effects, as increased mobility in areas with limited public transport tends to bring the average ridership lower. New investments are programmed to develop the transit network what has not been done for more than 50 years.

4 Data

Access to good quality data is an important concern in every modeling work. This is particularly the case for microsimulation models that are avid in large volume of disaggregated data. Access to individual detail data is costly in terms of their value and resources needed for their treatment. Moreover, it is also limited because of privacy protection legislations. In this case study the partnership with public agencies allowed us to access rich data sources with a particular interest to this study.

4.1 Data sources

Various data sources describing population, employment, land use, real-estate prices and transportation system are used in this study. They are presented below:

4.1.1 Population data: Census

The French Census has been reformed since 2004. Before, a general census was carried out every 10 years. The new census becomes yearly but only a sample of population is surveyed. Below we are describing the two Census types.

General Census of 1999 *(RGP99)*

The recent general census has been carried out in years 1982, 1990 and 1999. For year 1999, we have access to the exhaustive individuals database with many of useful variables but not all of them. This database constitutes our main source of population data in the Paris case study and in this case we did not need to create a synthetic population base. Based on this data, we estimate the residential location choice model and also the yearly rates of residential relocation.

Census data are also used to provide the thematic data tables for 13 subjects: Population (Age, sex, matrimonial status, activity, etc.), activity of residents (activity rates by sex, unemployment, employment conditions, etc.), employment (status as employees or self-employed, part-time or full-time, ...), commuting, education (students at their study location, diplomas, education level, ...), migration (birth place, residence *commune* at the time of the previous census), citizenship and immigration (French or foreign nationals, immigrant or French at birth, etc.), households (sex, age, education level and activity of the household head, size, number of active members, etc.), dwellings (type, number of rooms, ...), dwellings occupied as principal residence (date of move in, tenure, surface, etc.), buildings (date of construction, number of floors, number of housing units, etc.) and finally the information at the family level (couples, single parents, citizenship and socio-professional category of the family head, who may differ from household head in multi-family households, number and ages of children, etc.). The probability to relocate is computed using the 5% sample of whole country base (which is the only one available for us with this variable) to be able to take into account the households that move out of the region. The study of commuters behavior is also based on 5% sample to observe simultaneously residence and workplace information. Other models are estimated using the complete database. For the base year, the aggregated variables for different geographical units are computed using the census data table regarding either residents at residence place or jobs at workplace.

The yearly census from 2006

The results of yearly census are provided as processed tables (by National Institute of Statistics and Economic Studies, *INSEE*) and also at individual level (a sample including the variable "sampling rate" and a limited number of variables). We use that information for calibration and validation of models. As they are provided every year, they allow us to check the results against observed evolutions and trajectories.

4.1.2 Employment data

Regional Employment Survey *(Enquête Régionale sur l'Emploi, ERE)*

It constitutes our main data source about jobs and businesses (**?**). It provides us with a set of fairly detailed indicators about activities and employees in both public and private establishments over the region. The survey has been carried out every 3-4 years up to 2001. Then it was replaced with another statistical database to which we could not have access. We had access to these tables for 1997 and 2001. They have been matched based on the unique establishment ID number assigned by national statistical authority. The base provides the location of establishment at commune level and the number of employees as well as the activity sector in a 700 items list.

4.1.3 Land-use and Real-estate data

MOS *(Mode d'Occupation du Sol)*

It is a land cover database prepared by the regional urbanism institute *(IAU-IdF)* (**?**). **EvoluMOS** is a database which tracks the land-use evolution in the Ile-de-France region. It is a database prepared and updated by *IAU-Idf*. It is based on information on land use collected in 1982, 1987, 1990, 1999, 2003 and 2008 to monitor and analyze in great detail the evolution of land use on the whole area. *MOS* provides at given points in time the surface of land covered by 83 land use categories. Less detailed classifications are also used (based on 47, 21 or even 11 categories).

MOS and *EvoluMOS* together characterize the spatial and temporal evolution of land use in Ile-de-France. It is an essential source of data for the analysis of the choice made by urban developers concerning land use transformation. Currently, we use four generic types in order to keep the presentation manageable. These are: Single house, collective housings, business, industry and service activities and others. Our purpose is to capture the main trends characterizing land-use changes.

Real-estate prices *The Cote Callon*

The Cote Callon provides the price and the rent values per unit floorspace for communes throughout France. These data span over a period of 10 to 12 years and distinguish old and new buildings and dwellings by number of rooms and standing category. It includes residential, industrial and commercial buildings, offices and constructible and agricultural lands. Prices stored in the database are average ones per square meter at *commune* level. It should be noted that we don't take into account the diversity of price within the same *commune*.

Information on land use taxes

The information about the different taxes to be paid by firms and residents, residential owners and tenants are also available and used in the models.

Other data

Other data collected mainly on the website of *INSEE* have been used to take account of inflation, vacancy rate by *département*, ...

4.1.4 Transportation Data

The *DRIEA* (regional infrastructures authority) provides us with the data used in the transportation model *MODUS* (which is a standard 4-stage static model with public and private transportation).

- Information on road network:
  - The Origin - Destination (O-D) matrix for car equivalent vehicle for morning and evening peaks,
  - number of equivalent cars using each link of the network during morning and evening peak hours,
  - for the year 2009, travel time, speed and traveled distance for all different O-D pairs,
  - for the coming years the same information as discussed above is available.
- Public transportation data:
  - User matrix for morning peak,
  - loaded network for morning peak (information on the number of passengers on each link of the public transportation network is provided),
  - travel time for at morning peak for the different O-D pairs. Detail information is also available such as in-vehicle time, access time to the public transportation from home, walking time, waiting time and walking time to destination.

To these data, we add the dynamic behavioral parameters needed for running the dynamic transportation model, METROPOLIS. These parameters that describe the departure time choice of travelers are described in chapter 3.3. Their values are estimated based on a specialized survey named MADDIF (de Palma & Fontan 2001).

4.2 Data preparation

The raw data described in the previous section were usually not in a format which could be directly used in UrbanSim. The information contained in these data was also usually insufficiently detailed with respect to the need of the different modules of UrbanSim. Extrapolation was sometime needed.

4.2.1 Typology of households (Control Totals)

UrbanSim needs the total number of households in the whole of region as input. It depends on the scenario and on the year. We have used 24 different types of households, which are described in Table 1. The households are distributed according to their size (1, 2, 3, 4, 5 and 6 or more persons), to their dwelling type (houses or flats) and to the ownership status (owner or tenant). Note that the number of households in these 24 types has changed considerably during the last decade (the typical household size is smaller, the fraction of owners is larger, etc.).

4.2.2 Typology of jobs

To take account of the distribution of jobs by activity sector, it was necessary to split service and commercial activities in 11 categories as follows:

- agriculture
- industry
- construction
- commercial activities, retail
- transportation
- financial activities
- real-estate activities
- business services
- personal services
- education, health and social affairs
- administration.

N. of persons	Dwelling Type	Tenure Status	1999	2006	2008
1	house	own	3.80%	3.89%	3.97%
2	house	own	7.84%	8.06%	8.09%
3	house	own	4.79%	4.54%	4.45%
4	house	own	5.49%	5.26%	5.22%
5	house	own	2.09%	2.08%	2.07%
6 and more	house	own	0.83%	0.77%	0.76%
1	flat	own	9.21%	9.43%	9.58%
2	flat	own	6.80%	7.00%	7.02%
3	flat	own	3.13%	2.97%	3.04%
4	flat	own	2.33%	2.24%	2.33%
5	flat	own	0.71%	0.70%	0.72%
6 and more	flat	own	0.27%	0.25%	0.25%
1	house	rent	0.71%	0.72%	0.72%
2	house	rent	0.80%	0.82%	0.81%
3	house	rent	0.67%	0.64%	0.63%
4	house	rent	0.65%	0.62%	0.60%
5	house	rent	0.29%	0.29%	0.29%
6 and more	house	rent	0.17%	0.16%	0.16%
1	flat	rent	20.90%	21.39%	21.35%
2	flat	rent	12.47%	12.83%	12.71%
3	flat	rent	7.11%	6.74%	6.73%
4	flat	rent	5.09%	4.88%	4.83%
5	flat	rent	2.28%	2.27%	2.29%
6 and more	flat	rent	1.56%	1.45%	1.39%

Table 1: Typology of households in different census years (Source: Census data, *INSEE*)

In order to take into account the wide range of behaviors (in terms of birth, death, growth and location) of firms as a function of their size, firms were classified in 2 wide categories of more or less than 10 employees.

4.2.3 Real-estate prices

From data available on real-estate prices in the *Cote Callon*, 5 types of real-estate were defined: Flats for rent or for sale, houses for rent or for sale and offices for rent. All prices are expressed in euro per square meter. The raw data obtained from *Cote Callon* are much more detailed and need to be grouped. The raw data differentiate the dwellings according to the age, comfort level, number of rooms, etc. The level of detail varies from one *commune* to another. For example, the partition according to the number of rooms is more detailed in the historical Paris than elsewhere in Ile-de-France.

To compute the average price value for the 5 types of real-estate properties, we have evaluated for each *commune* the weighted average of the more detailed prices. The weights were inferred from Census data, which provide information for example on the number of rooms in the dwellings or on the construction age.

The prices of the *Cote Callon* were available for *communes* with more than 5000 inhabitants. To obtain the price for other *communes*, we have estimated the hedonic model of price on the sample, where they were observed. We then used such estimation to predict the price for all the *communes* in Ile-de-France.

4.2.4 Land use

For the land use transition module, the different types of land uses are classified into 4 general types: Single houses, Flats or Collective dwellings, Activities and others (see Table 3). These 4 categories capture the guidelines for the most relevant land use transition and to analyze and predict the urban development. Parallel to this aggregation process of the different types of use, we have studied to which extent each type of land use could accommodate different types of business or governmental activities. Some combinations have a very low probability to occur and on the contrary some types such as green areas and highly accessible places can only accommodate a limited number of activities. The institutional regulation, zoning and fiscal constraints have also been taken into account.

5 Data analysis: Spatial issues

5.1 Spatial issues

The area can be split in 3 rings, 8 *départements*, 44 *arrondissements*, 1 300 *communes*, 5 249 *IRIS* and about 50 000 grid cells (500 meters side). Table 5 presents accessibility and environmental variables and indicates their average values in each *département*. The attributes are weighted by the number of households in each entity in order to compute these averages. Many statistics are only available at *commune* level. Table 5 indicates average value of these variables.

Table 6 shows the variance decomposition at the three relevant geographical levels: *Département*, *commune* and grid cell. For each variable X (such as green area or accessibility, etc.), the model is:

$$X_{dcg} = \mu + \eta_d + \xi_c + \varepsilon_g, \tag{1}$$

where, according to the analysis of variance model, X_{dcg} denotes the value of X in grid cell g located in *commune* c and in district d, μ is the average level of X in Ile-de-France, η_d, ξ_c and ε_g are independent random Gaussian variables with zero mean and variances σ_d^2, σ_c^2 and σ_g^2, respectively. As the polygons are independent

Code	**Label 47 types**	**4 types**	**Poten. to activ.**	
			SR	**MR**
1	Forests	Others		
2	Clearings in forest	Others		
3	Poplar plantation	Others		
4	Farm lands	Activities		20%
5	Field with grass for agriculture	Activities		20%
6	Orchards, tree nurseries	Activities		20%
7	Gardening, horticulture	Activities		20%
8	Intensive greenhouse cultivation	Activities		20%
9	Water	Others		
10	Non-agricultural surfaces with grass	Others		100%
11	Quarries and pits	Activities		
12	Landfills	Activities		100%
13	Rural vacant space	Others		100%
14	Parks or gardens	Activities		
15	Allotments	Houses		
16	House gardens	Houses		
17	Outdoor sport field	Activities		
18	Large sport equipments	Activities		
19	Camping, caravaning	Activities		
20	Related to recreation park	Activities		
21	Vacant urban land	Others		100%
22	Individual housings	Houses		
23	set of Identical individual housings	Houses		
24	Rural housing	Houses		
25	Low-rise housings (not fragmented)	Houses		
26	High-rise collective housings (not fragmented)	Flats	20%	20%
27	Fragmented collective housings	Flats	10%	10%
28	Other housings	Flats	20%	20%
29	Water, sanitation and energy equipments	Activities	100%	100%
30	Spaces allocated to Activities	Activities	100%	100%
31	Logistic warehouses	Activities	100%	100%
32	Commercial spaces	Activities	100%	100%
33	Offices	Activities	100%	100%
34	Sport facilities	Activities		
35	Educational facilities	Activities	100%	100%

Table 3: Land-Use categories and the 4 general types, short run and medium run potential to accommodate activities (Source: *IAU-IdF* and authors' appreciation)

Code	Label 47 types	4 types	Poten. to activ. SR	MR
36	Health facilities	Activities	100%	100%
37	Cemeteries	Activities		
38	Large convention centers and exhibition	Activities		
39	Cultural and leisure facilities	Activities		
40	Governmental buildings	Activities	100%	100%
41	Other public facilities	Activities	100%	100%
42	Railway transportation	Activities		
43	Roads	Activities		
44	Parking lots	Activities		
45	Bus stations, depots	Activities		
46	Airports	Activities		
47	Construction sites	Activities		

Table 3: cont.

in all levels, there is no correlation and covariance and we have the total variance of X denoted by $\sigma^2 \equiv \sigma_d^2 + \sigma_c^2 + \sigma_g^2$, and the percentage of variance (reported in Table 6) at level i, $i = d, c, g$, is:

$$\alpha_i = 100 * \sigma_i^2 / \sigma^2,$$

with $\alpha_d + \alpha_c + \alpha_g = 100$.

Note that the term redevelopment area used in Table 5 is not a precise translation for the term '*Zones Urbaines Sensibles*'. These are areas with high concentrations of social and economic difficulty. They are targeted for government assistance or redevelopment investments. We will use the term Redevelopment Areas in this context.

We observe that for many of variables observable at an infra-communal level, a great part of variation is observed at that level finer than commune. That suggests to use an infra-communal geographical unit for modeling. But many important variables are not available at such a level. We denote also some communes with very low number of inhabitants and jobs (lower than 100) that are essentially situated at the far suburbs. We may suggest a multilevel (nested) geographical unit for modeling to take into account all available information. In this chapter we report the results of simulations done at commune level that was sufficient for the precision required for our policy analyses.

Variable \ *département*	**75**	**92**	**94**	**93**	**78**	**91**	**95**	**77**
Number of Grid Cells	420	704	980	952	9,400	7,399	5,187	24,194
Number of *communes*	20	36	47	40	262	196	185	514
Distance and accessibility								
Subway and tramway stations	8.17	0.80	0.65	0.72	0.00	0.00	0.00	0.00
Subway and tramway stations around	71.93	7.14	5.77	6.49	0.00	0.00	0.00	0.00
Train stations	0.98	0.75	0.44	0.53	0.53	0.47	0.68	0.30
Train stations around	9.35	6.89	3.94	4.76	4.29	4.07	5.60	2.35
Accessibility to jobs (Public Transit)	50.93	49.29	48.02	48.25	45.20	44.30	45.97	41.86
Accessibility to jobs (Private Car)	54.45	53.88	53.42	53.48	51.18	51.15	51.82	48.60
Accessibility to jobs[5]	*-72.9*	*-72.3*	*-73.9*	*-74.0*	*-73.8*	*-93.3*	*-68.8*	*-89.9*
Average travel time (Public)	*28.08*	*31.54*	*38.45*	*38.54*	*47.84*	*52.04*	*46.24*	*48.86*
Average travel time (Private)	*16.19*	*16.08*	*16.31*	*16.90*	*24.13*	*36.58*	*20.31*	*35.59*
Accessibility to shops (Public)	33.24	32.25	31.98	32.06	30.28	30.25	30.94	28.66
Accessibility to shops (Private)	35.58	35.30	35.30	35.31	34.06	34.27	34.48	32.96
Private car travel time variability	*2.72*	*4.11*	*3.07*	*3.46*	*3.55*	*6.65*	*3.71*	*3.56*
Distance to arterial	2.03	1.28	0.74	0.67	3.27	1.62	2.71	2.15
Distance to highway	1.48	1.19	1.27	1.28	2.78	3.12	2.34	4.68
Distance to Châtelet	3.57	8.95	11.42	10.88	27.64	26.36	20.27	41.50
Environmental variable								
% surface in noisy areas (severe noise: >96dB)	0.00	0.00	4.44	0.03	0.00	2.32	1.36	0.76
% surface of parks and gardens	7.59	9.13	9.54	8.65	13.09	12.66	10.60	12.74
% surface covered by water	0.90	1.37	1.80	0.65	1.42	0.98	0.71	1.43
% surface covered by forests	0.27	1.54	1.20	1.01	7.15	7.28	3.97	8.02
% surface: Sport facilities	1.43	3.05	2.48	3.03	2.42	2.36	2.89	1.77
% surface: Public facilities	9.32	13.72	13.20	12.69	22.67	22.31	17.46	22.52
% surface: Open space	8.95	12.13	12.60	12.54	16.18	15.65	14.31	14.99
% surface: Redevelopment area	5.01	8.32	5.23	8.14	2.18	6.67	6.57	1.63

Table 5: Average accessibility and environmental variables, by *département* (over the grid cells and weighted by population). Variables in italics are computed at the *commune* level and not represented in Table 6 (Source: authors' computations from *IAU-IdF* GIS database and METROPOLIS computations)

Variable	*département*	***Commune***	**Cell**
Distance and accessibility			
Number of subway and tramway stations	60.5%	22.1%	17.4%
Number of subway and tramway stations around	65.2%	22.3%	12.4%
Number of train stations	10.0%	23.1%	66.9%
Number of train stations around	17.0%	33.4%	49.6%
Accessibility to jobs (Public Transit)	35.1%	64.9%	0.0%
Accessibility to jobs (Private Car)	39.1%	60.9%	0.0%
Accessibility to shops (Public Transit)	34.9%	65.1%	0.0%
Accessibility to shops (Private Car)	35.8%	64.2%	0.0%
Distance to nearest arterial	7.8%	84.1%	8.1%
Distance to nearest highway	7.8%	83.0%	9.2%
Distance to Châtelet (Paris center)	42.7%	56.9%	0.4%
Environment			
% surface in Noisy area (severe noise: >96dB)	1.8%	62.4%	35.9%
% surface of Parks and Gardens	5.2%	17.2%	77.6%
% surface of Water	7.4%	19.3%	73.4%
% surface of forests	2.5%	36.2%	61.3%
% surface in ‘Redevelopment Areas’	3.9%	17.3%	78.8%
% surface of Public Facilities	0.5%	19.3%	80.3%
% surface of Open Space	1.9%	37.6%	60.5%
% surface of Sport Facilities	2.0%	11.3%	86.7%

Table 6: Variance decomposition of accessibility and environmental variables over different scales (Source: authors’ computations from *IAU-IdF* GIS database and METROPOLIS computations)

5.2 Households moving rate

Residential moving (relocation) strongly depends on the household types as can be seen in Table 7.

	Household size					
Over 1 year (1998)	**1**	**2**	**3**	**4**	**5**	**6+**
house, own	4.94%	5.73%	7.06%	7.73%	7.50%	7.32%
flat, own	7.13%	6.35%	7.72%	6.60%	6.49%	7.68%
house, rent	22.20%	24.99%	25.51%	25.41%	23.21%	18.86%
flat, rent	19.18%	17.33%	15.08%	12.29%	11.65%	8.67%
all	13.98%	11.15%	11.60%	10.35%	10.19%	8.85%
Over 9 years (1990-1999)	**1**	**2**	**3**	**4**	**5**	**6+**
house, own	32.32%	35.71%	44.78%	57.93%	58.33%	52.90%
flat, own	41.46%	36.28%	49.73%	56.48%	58.08%	53.09%
house, rent	74.18%	76.31%	80.92%	85.55%	85.34%	78.86%
flat, rent	62.79%	59.85%	65.09%	67.22%	64.18%	54.83%

Table 7: Households moving rate, by household type (Source: Census data, *INSEE*)

Globally the relocation rate decreases with the household size. It should be noted that the impact of household size also depends on tenure and type of housing. Tenants are generally more mobile than owners. Owners move more often when they initially own a flat. By contrast, tenants move more when they initially rent a house. These results are consistent with the fact that households either buy a house that they keep it several years (still, household competition has changed significantly) or rent a flat and move frequently in response to professional job change of the different members of the household.

5.3 Working zoning

In the context of Grand Paris, the results of model Relu-Tran (Anas & Liu 2007) based on about 50 zones have been compared with those of UrbanSim based on 1 300 zones. The definition of these zones depend on administrative boundaries on one hand and on constraints imposed by Grand Paris on the other hand. In the current form, the zones depend on the poles of economic development. They are 10 poles of economic development also referred to as *Contrats de Développement Territoriaux (CDT)*. Various difficulties arose during the process of defining the boundaries of the zones. Indeed several CDTs include part of several *départements* or *communes*. The final solution which was selected is a compromise between the logic behind the definition of CDT and the logic behind the zones used in the transportation model. Figure 5 presents the 50 zones selected which include 20 *arrondissements* within Paris and 10 zones which match approximately the *CDTs*.

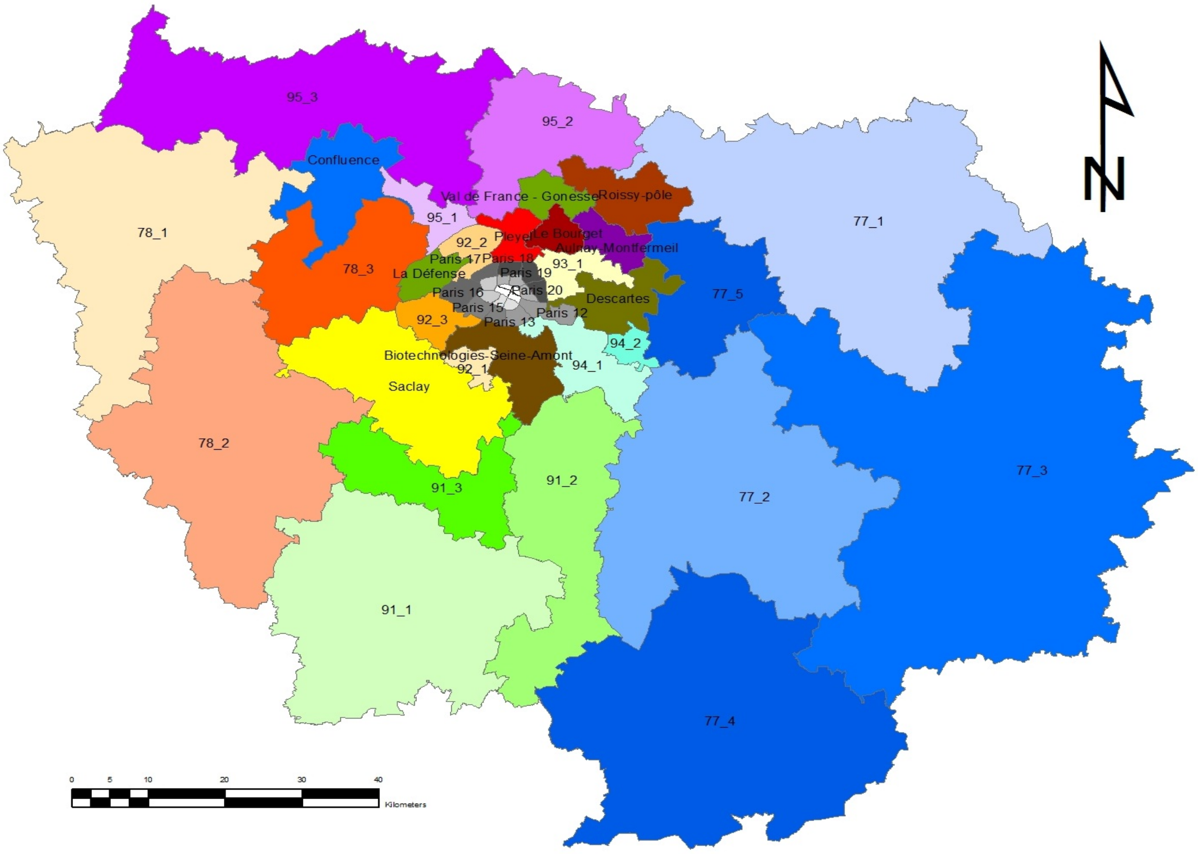


Figure 5: Aggregate zones defined in collaboration with *Société du Grand Paris*

6 Model structure

UrbanSim is constituted by several models. The urban development model describes the endogenous land use change as an outcome of a dynamic process which is driven by the developers. The stakeholders (developers, public authorities, etc.) react themselves to satisfy the expected global demand of the households and of the firms (see Chapter of 2.8 on this issue). A development project that consumes some surface of land and concerns a given number of dwellings or a given activity floor space. The project will be located in a parcel, consistent with regulation, which maximizes the profit of developer. For example, the probability that a given project of dwelling or activity floor space be located in a given parcel depends on the local prices as well as the demand for housing and commercial activities. This process provides a quite realistic description of the competitive use of land. Of course, competition remains imperfect and some political economy dimension should be included in the future work to account for this deviation from perfect competition.

The potential land use are defined exogenously. They depend on environmental regulation (e.g. preservation of forests) and on constraints imposed by the local urban plan *(Plan Local d'Urbanisme)*. Time lags are important. For example there is a time lag of one year between the decision to allocate land to some residence or commercial activity and the actual available supply. In other words the supply of housing and commercial unit is inelastic in the short run (1 year), but elastic in the medium run. It is assumed that, on the demand side, agents are price takers since they are small and therefore unable to have an impact on the price level (this is an assumption used in perfect competition studies). The price adjusts for one year to another as a function of supply and demand in each of the sub markets (e.g. flats, houses, purchase or the rent of dwellings and offices for rent). Households and firms behavior depends on various factors such as prices, accessibility and local amenities. Their main behavior modeled is the location choice.

Location choice modeling

The location choice of jobs and workers is determined by the location of the firm and of the households. Residential location is specific to each type of housing. There are several determinants of residential location. For example density, accessibility to job and shopping areas, local supply of dwellings, income of households of the vicinity, local amenities and the proximity to the city center. It is measured by the distance to Chatelet (the point of highest accessibility for public transportation) and to expressways. UrbanSim assumes inelastic supply of the dwelling in short run (over 1 year). The choice of residential location is a first come first served principle; households are sorted in a random way. At the beginning of the year the given number of locations are available at each location (*commune*). This determines the choice set of the first household which wish to find a residential location. The choice set of the next household in the list is the same as the previous one except that one vacant unit is now occupied and therefore unavailable. Housing prices are updated on a yearly basis but do not vary within a given year. Since the prices are fixed in the short run (one year of period) there may be rationing of some housing. This process is compound by the fact that the markets are fragmented.

Employment location depends on the accessibility with respect to the worker, on agglomeration force, complementarity and substitutability between business activity sectors, on local population density, on the local distribution of land use, on the distance to the road infrastructures (streets, roads and motorways) and on the distance to the center of the region.

6.1 Household models

6.1.1 Demographic model

The demographic module is described in Chapter 2.1. A microsimulation model was developed by *INED*[6] and allowed to create and update the yearly population database for the period 1999 - 2050.

6.1.2 Household Residential Location Choice Model

We have investigated what is the most appropriate spatial scale for studying preferences for locational amenities. This raises the question of the most relevant scale for estimating a household location choice model. We were asking whether residents have preferences for *communes* or for smaller more homogeneous geographical units. In order to answer this question, we estimated two location choice models: one at the *commune* level, and the other one at the grid cell level.

The *commune* j contains C_j dwellings, j=$1, \ldots, 1300$. We assume that all the dwellings i located in *commune* j have the same observable attributes, and therefore the same expected utility $V^h_{i\in[1,C_j]} = V^h_j$ for household h, h=$1, \ldots, N$. The total number of dwellings in Ile-de-France is denoted by I. Typically the value of I is about five million dwellings over the area.

The probability that household h chooses a dwelling i located in *commune* j is given by the Multinomial Logit formula:

$$\mathrm{P}^h_i = \frac{\exp\left(V^h_i\right)}{\sum_{i'=1}^{I} \exp\left(V^h_{i'}\right)}. \tag{2}$$

Since we have assumed that all the dwellings located in j have the same expected utility $V^h_{i'}$, they therefore have the same probability of being selected $P^h_{j'}$, Equation (2) implies that the probability that household h selects *commune* j is:

$$\mathrm{P}^h_j = C_j \mathrm{P}^h_i = \frac{C_j \exp\left(V^h_j\right)}{\sum\limits_{j'=1}^{J} \left(\sum_{i'\in j'} \exp\left(V^h_{i'}\right)\right)} = \frac{\exp\left(V^h_j + \log\left(C_j\right)\right)}{\sum\limits_{j'=1}^{J} \exp\left(V^h_{j'} + \log\left(C_{j'}\right)\right)}. \tag{3}$$

Under the IIA (Independence of Irrelevant Alternatives) assumption, one can obtain consistent estimates of the parameters in the expected utility function by selecting a random sample of alternatives, with uniform sampling of alternatives, provided the correcting term $\log\left(C_j\right)$ is added to the likelihood. However, more efficient estimates can be obtained with the technique of importance sampling of alternatives, that is if the probability that alternative j is included in the choice set is proportional to C_j, provided a second correcting term $-\log\left(C_j\right)$ is added to the likelihood. It can be shown that no correcting factor is necessary to obtain consistent estimates

[6] *Institut national d'études démographiques*

of the parameter when importance sampling is used. (See Ben-Akiva & Lerman (1985) and de Palma et al. (2005)) for details on this issue.

For the model at grid cell level, assuming that all the dwellings i located in the grid cell k, k =1,...,K, which is located in *commune* j, j *=1,...,J,* have the same observable attributes, household h, h *=1,...,N,* have the same expected utility $V_i^h = V_k^h$ for them. We assume that the expected utility is a linear combination of grid cell attributes X_k and *commune* attributes Z_j, in which the marginal utilities of grid cell and *commune* attributes can be household-specific. Expected utility is therefore of the form:

$$V_i^h = X_k\beta_h + Z_j\gamma_h, k \in j, \tag{4}$$

where β_h and γ_h denote the household-specific marginal utilities of grid cell and *commune* attributes, respectively.

If households choose among *communes* rather than grid cells, then the expected utility of a dwelling located in *commune* j only depends on Z_j and is of the form:

$$V_j^h = Z_j\delta_h = Z_j\gamma_h + E\left(X_k\beta_h|Z_j\right), \tag{5}$$

where δ_h mixes the marginal utilities of *commune* attributes and the marginal utilities of grid cells attributes, weighted by their distribution in the *commune*. The exact formula for $E\left(X_k\beta_h|Z_j\right)$ corresponds to the log-sum in a nested logit model (see Anderson et al. (1992) for details).

The number of dwellings in *commune* j and in grid cell k are denoted by C_j and G_k, respectively. Since all the dwellings located in k have the same expected utility, and the same probability of being selected, Equation (3) implies that the probability that household h selects grid cell k is:

$$\mathrm{P}_k^h = G_k\mathrm{P}_i^h = \frac{G_k\exp\left(V_k^h\right)}{\sum_{k'=1}^{K}\sum_{i'\in k'}\exp\left(V_{i'}^h\right)} = \frac{\exp\left(V_k^h + \log\left(G_k\right)\right)}{\sum_{k'=1}^{K}\exp\left(V_{k'}^h + \log\left(G_{k'}\right)\right)}. \tag{6}$$

Similarly, the probability that household h selects *commune* j is:

$$\mathrm{P}_j^h = C_j\mathrm{P}_i^h = \frac{C_j\exp\left(V_j^h\right)}{\sum_{j'=1}^{J}\sum_{i'\in j'}\exp\left(V_{i'}^h\right)} = \frac{\exp\left(V_j^h + \log\left(C_j\right)\right)}{\sum_{k'=1}^{K}\exp\left(V_{j'}^h + \log\left(C_{j'}\right)\right)}. \tag{7}$$

We report in Table 9 the estimation results for the household location choice models at the level of 500 x 500 meter cells within the study area, and at the level of *communes*.

6.2 Employments models

In Paris case study, we have estimated and implemented two versions of employment location models. The first one was a job location model and the second a firmographic model (**?**). Firmography provides a more realistic and comprehensive knowledge of the evolution of employment over the studied area and also about the establishments.

Variable	***Commune* level**		**Gridcell level**	
	Coeff	**t-stat**	**Coeff**	**t-stat**
Same district as before move	2.539	275.9	2.535	272.5
Paris	-0.254	-10.10	-0.168	-10.29
Log(Price of flat)	0.006	0.20	0.021	0.94
Log(Price of flat)* (Age-40)/10	-0.029	-1.74		
Log(Price of flat)* (Log(Income)- $\overline{\text{Log(Income)}}$)	0.409	8.20	0.248	5.84
Log(Price of flat)* (N. of hh members – 1)	-0.042	-2.40		
Log(Price of house)	0.087	3.70	0.139	7.41
Log(Price of house)* (Age-40)/10	-0.104	-8.78	-0.114	-20.90
Log(Price of house)* (Log(Income)- $\overline{\text{Log(Income)}}$)	0.187	5.78	0.149	5.17
Log(Price of house)* (N. of hh members – 1)	-0.109	-8.38	-0.130	-20.43
N. Subway stations around			-0.004	-2.37
N. Subway stations in the *commune* / cell	0.005	4.59	0.001	0.21
N. Railway stations around			0.013	3.85
N. Railway stations in the *commune* / cell	-0.009	-4.86	0.006	0.60
Average travel time from j, commuting (TC) [hr]	0.025	0.80	0.040	1.45
TC*(Dummy female) [hr]	-0.374	-8.28	-0.294	-6.56
Distance to highway [km]	-0.005	-3.10	-0.001	-0.74
Distance to arterial [km]	-0.008	-2.98	-0.015	-5.87
Distance to Châtelet (Paris centre) [km]	0.002	2.94	-0.001	-1.06
% households with 1 member * 1 member in h	1.877	20.62	2.270	30.73
% hhs with 2 members* 2 members in h	1.336	4.20	1.773	10.03
% hhs with 3+ members* 3+ member in h	2.230	22.16	2.075	27.98
% hh with no working member * no working member in h	7.797	33.91	5.382	33.49
% hh with 1 working member * 1 working member in h	-0.931	-5.84	0.502	4.58
% hh with 2+ working member * 2+ working member in h	1.664	14.83	0.404	5.01
% hh with a young head	0.450	3.01	1.021	10.86
% hh with a young head * young head in h	4.777	28.34	3.243	29.59
% hh with a middle age head * middle aged head in h	-0.693	-4.16	-0.162	-1.51
% Rich hh * Rich h	3.280	28.36	2.876	36.58
% Medium Income hh * medium income h	1.621	10.11	1.923	17.00
% Poor hh * poor h	0.459	3.53	1.070	12.11
% hhs with a foreign head * foreign head in h	5.046	26.74	4.578	37.26
% hhs with a foreign head * French head in h	-1.995	-17.44	-1.305	-16.71
% of surface in Redevelopment Area * Rich h	0.678	4.97	-0.228	-4.28
% of surface in Redevelopment Area * Med. Inc. h	-0.083	-0.78	-0.193	-5.16
% of surface in Redevelopment Area * Poor h	0.431	4.45	0.165	4.99

Table 9: Household location choice: estimation results at *commune* and grid cell levels (Source: authors' estimations)

Variable	*Commune* level		Gridcell level	
	Coeff	t-stat	Coeff	t-stat
Log of the number of residential units	0.067	9.26	0.023	3.27
% of flats in total dwellings * Foreign head in h	1.368	15.42	0.676	11.37
% of flats in total dwellings * French head in h	0.485	7.28	0.161	3.76
% of flats in total dwellings * (N. of members - 1)	-0.085	-4.29	-0.056	-4.29
% of flats in total dwellings * young head in h	-0.131	-2.43	-0.037	-1.01
% of flats in total dwellings * old head in h	-0.766	-8.40	-0.188	-2.82
% of surface in Noisy area	-0.089	-1.56	-0.077	-1.52
% of surface covered by Forest	-0.117	-3.09	-0.297	-5.05
% of surf. covered by Forest * N. of Children	0.423	6.87	0.461	4.99
% of surface covered by Water	0.204	1.62	0.333	4.24
% of surface covered by Parks and Gardens	-0.428	-4.22	-0.152	-2.93
% of surf. covered by Parks * N. of children	-0.129	-0.71	0.615	6.67
% of surface covered by Sporting facilities	0.442	2.50	-0.371	-4.06
% of surf. covered by Sporting facilities * N of children	0.268	0.83	0.786	4.93

Table 9: Cont.

6.2.1 Background and definition

The term firmography, by analogy with demography, can be assigned to a fourfold model which includes establishment birth, its re/location, variation in the number of jobs within establishment, and its ultimate destruction. Consistently with available data, we propose some theoretical models which describe three stages of the establishments' life cycle: Disappearance of the establishments, evolution of the workforce within the stable units and location of new establishments. For more discussion of the life-cycle of firms we refer the reader to the Chapter 3.1.

To get an overview on the first stage of the establishment life cycle, the birth of the establishment, we compute the creation rates across eleven activity sectors for each county of Ile-de-France (details are provided in SustainCity Working Paper 3.7). All models are estimated separately for each activity sector.Given the volume of results, we refer the reader to SustainCity W.P. 3.7 for more details. We will focus here on three key models: The first one describes establishment disappearance; the second one describes the evolution of the establishment workforce and the third one describes the location choice, Finally we briefly discuss the implementation of firmographic models in UrbanSim.

Model of Establishment Disappearance

We model the establishments death using a binary probit model. The death of the establishment is registered in two cases: When the establishment is closed, but also when it relocates. Consistently with the available data, we are not able to track the establishments which move. For this reason, we do not model the relocation of the establishments. When the establishment moves, as its id[7] is automatically changed so that there is no possibility for us to identify the "same" establishment before and after the move. Then, we consider this event as a death and a birth of a new establishment. For an establishment present in 1997, the probability to disappear is:

$$P\left(e \notin E^{2001}\right) = P\left(VL_e > 0\right) = P\left(\beta_s X_c + \gamma_s Y_e + \varepsilon_e > 0\right),$$

with

$$VL_e = \beta_s X_c + \gamma_s Y_e + \varepsilon_e,$$

where e - establishment present in the market in 1997

E^{2001} - set of establishments present in 2001

VL_e - value of leaving the market by the establishment e

ε_e - random part which measures the fragility of the establishment; $\varepsilon_e \sim N(0,1)$

X_s - *commune* characteristics

β_s - vector of coefficients which stands by the *commune* variables

Y_e - establishment characteristics

γ_s - vector of coefficients which stands by the establishments characteristics

$P\left(e \notin E^{2001}\right)$ - probability that the establishment is not in the market any more in 2001.

We check whether the following variables have a significant effect on the disappearance of the establishment:

- the initial size of the establishment,
- accessibility to population with particular socio-economic characteristics,
- regional accessibility measures, such as number of metro stations, average time for public transport or private vehicles,
- accessibility to labor (workforce from different socio-professional classes, workforce with different education levels) or
- real estate prices.

Model of Establishment Evolution

To model the workforce evolution of the stable establishments, i.e. the establishments which are observed in the market for the two analyzed periods, we use regression model with the logarithmic value of the relative difference between establishments size in the final period and the initial year as a dependent variable. We first consider

$$\ln\left(\frac{\tau_e^{2001}}{\tau_e^{1997}}\right) = const_s + \sum_{t \in T} \beta_{s,t} I_t + \sum_{t \in T} \gamma_{s,t} I_t \left(\tau_e^{1997} - \tau_t^{\inf}\right) + \delta_s X_c + \varepsilon_e.$$

Alternatively, we consider another specification

$$\ln\left(\tau_e^{2001}\right) = const_s + \sum_{t \in T} \beta_{s,t} I_t + \sum_{t \in T} \gamma_{s,t} I_t \left(\tau_e^{1997} - \tau_t^{\inf}\right) + \delta_s X_c + \varepsilon_e.$$

[7]The establishment has a unique id number assigned by the *INSEE*. It is named SIRET code.

The evolution model describes the growing or shrinking behavior of the establishment. Final workforce is a function of the initial workforce and the variables which represent the socioeconomic characteristics of the population and employees living in the *commune* surrounding of the establishment.

We use the following notation:

I_t - size class indicator

t- size class

τ_e^{2001} -workforce (size) of the establishment in 2001

τ_e^{1997} - workforce (size) of the establishment in 1997

$\tau_t^{\inf}$ - lower limit of each size class t

X_c - *commune* characteristics

δ_s - vector of coefficients which stand by the *commune* variables

ε_e - random perturbation.

To estimate the parameters in the evolution models, similar variables are used as in the disappearance models presented in the previous section.

Establishment Location Choice model

We use the multinomial logit model (McFadden et al. 1974 and McFadden 1981) to study the location choice of the business units. The probability that a *commune* is included in the choice set is proportional to the total number of employees in this particular *commune*.

$$P_i^e = \frac{\exp\left(V_i^e\right)}{\sum\limits_{i'=1}^{I} \exp\left(V_{i'}^e\right)},$$

and

$$P_j^e = C_j P_i^e = C_j \frac{\exp\left(V_i^e\right)}{\sum\limits_{j'=1}^{J} \sum\limits_{i' \in j'} \exp\left(V_{i'}^e\right)} = \frac{C_j \exp\left(V_i^e\right)}{\sum\limits_{j'=1}^{J} C_{j'} \exp\left(V_{i'}^e\right)} = \frac{\exp\left(V_i^e + \ln C_j\right)}{\sum\limits_{j'=1}^{J} \exp\left(V_{i'}^e + \ln C_{j'}\right)}.$$

As the residential location choice model, we assume that all locations which belong to *commune* j have the same observable characteristics and lead to the same expected utility level. Thus all the establishments in the same *commune* j have the same probability P_j^e to be chosen by the establishment. These models address location behavior of units at the establishment level. We model only the location choice of new establishments. As was mentioned before, according to the way the employment database is built, there is no possibility to distinguish between the real disappearance of the establishment and its relocation. To avoid correlation between variables, de Bok & Sanders (2005) propose to construct a variety of *composed* accessibility that describe the distances to the physical infrastructure, the accessibility to labor, the accessibility to the customers and suppliers. They also suggest using a set of agglomeration variables to explore concentration or deconcentration patterns. They use the information on the urban environment and they treat an average rental level in a particular territory as a proxy for the real estate quality. All the models presented by de Bok and Sanders are computed by establishments size and by activity sector. Waddell & Ulfarsson (2003) confirm the significance of the real estate characteristics, the neighborhood characteristics, local and regional accessibility measures (access to jobs, population and distance to highway, airport or central business districts) for the location models. Picard & Antoniou (2011) claim that commuting time is the most significant variable in explaining the location, much more important than variables which measure either accessibility or expected time commonly used in the location choice models.

In the location choice model, which we run, we take into consideration ten out of the 11 activity sectors (Agriculture sector is excluded[8]) and we test the significance of the following variables:

- the local competition in the employment market and the employment density in each activity sector,
- local and regional accessibility measures (variables indicating the distance to the national road, highways, travel time data, etc.),
- general characteristics of the region in which the establishment is being located,
- accessibility to population (financial situation of the households, the age of the head of the household, the presence of children in the household),
- real estate prices, treated as a proxy for the attractiveness of the area,
- land-use data.

Firmographics implementation in UrbanSim

The three models described above have been implemented in UrbanSim by instantiation of the existing generic linear regression and multinomial logit models of OPUS core (see **?**). The estimated models represent the evolution over four years between 1997 and 2001. Also, the models are describing the establishment behavior as a function of the urban condition 4 years before. To obtain realistic results, at each year the model uses 4 years old information and not the information of the last simulated year as it is the normal procedure of UrbanSim. The number of very large establishments is low and their behavior is different from other establishments. As a consequence, we have excluded the establishments with a workforce greater than 250 employees from our models. Any information about their evolution has been introduced as exogenous events. The control data that drive the simulation are given as the total number of jobs at each year differentiated by the activity sector. To guarantee the equality of the total number of jobs and the sum of the jobs in

[8]The number of observed creations or moves of establishment in this sector was too low to be able to estimate a significant location choice model. That will be the case in the simulated period, too. This sector will be treated exogenously.

active establishments, we have added a 4th module that along with the employment transition model controls for the total number of jobs. If in a sector and for the simulated year, the total number of jobs exceeds the target one, the transition model will insure that the appropriate number of establishments disappear in order to meet the target value. If after the workforce evolution step, the number of jobs over the region is lower than the target control value, the establishment reappearance model will reactivate the appropriate number of previously disappeared establishments.

6.3 Real-estate Price Models

The property prices are modeled using hedonic regression (**?**). We have performed simultaneous regressions for the price (rental price and purchase price) of offices, houses and flats. The method of hedonic regression models the interaction between supply and demand. It allows to compute the implicit valuation of the characteristics of the dwellings as well as of the characteristics exogenous to this dwelling including several local amenities identified in the literature. Moreover, the price of the specific housing units varies over time as a response to fluctuation between supply and demand. The difference between short term vacancy rate (for given year) and long term vacancy rate plays also an important role in these evaluation. For example, when the short run vacancy rate is below the structural vacancy rate, price level is triggered upward, and vice versa for the other situation.

The prices are updated each year after the construction and the market transactions occurred last year. The prices at the end of a given year are used as the reference prices for the decisions occurring in the following year.

6.4 Urban Development Models

The urban development model simulates the construction of new housing projects which follow new infrastructure development or to transformation of existing projects. The necessary data to estimate the coefficients for urban development models, rely on data of *MOS*. We identify the changes between two consecutive dates in the land use. The model predicts the probability that a project with a given land use type be located at a given *commune*. A multinomial logit model is used to estimate that. This probability is then estimated and a Monte-Carlo simulation is used to select the specific location of new project. More precisely, each simulation year the model computes the difference between total spaces and occupied ones. If the vacancy rate is lower than some critical value, new projects will be created accordingly. The model will determine exactly where this project will be located. The urban development model has three modules:

- the transition module which determine the needs for construction of housing units and which computes the quantities to be built (number of dwellings and business units). This module also selects the projects that are required to be performed;

- the module which computes the geographical location of the project;
- the module which updates the stock of buildings which in particular takes into account new constructions.

During the simulation, the model performs sequentially the following steps:

- Computation of the real vacancy rate (number of housing units minus number of households);
- comparison of the above vacancy rate with the structural vacancy rate;
- decision to build;
- computation of the surface to be built;
- selection of the project of a given size required to meet specific needs;
- computation the choice set (the *communes* which can potentially accommodate the specific project on the basis of the nature of the land supply);
- location of the projects;
- update of the buildings capacity.

7 Transportation Model

METROPOLIS is a fully dynamic transportation model that is specially adapted for large networks. It is a mesoscopic event based model and uses a multi-agent methodology with a disaggregated representation of travelers. The supply system relies on a macroscopic formulation that computes travel time as a function of the flow condition of the link. It has been developed since the 90’s and its main application on the Paris region has been within the QUATUOR project (THEMA-TT-R 2002). It models the mode, departure time and route choices of users. The Logit formula is used for these models. The dynamic assignment procedure can be deterministic or stochastic.
The generalized cost function is:

$$C\left(t\right)=\alpha tt^{VP}\left(t\right)+\beta\left[t+tt\left(t\right)-t^{*}\right]^{+}+\gamma\left[t^{*}-t-tt\left(t\right)\right]^{+}.$$

The generalized cost $C(.)$ includes the schedule delay cost terms in $(t^{*}-t_{a})$, where t^{*} is the desired arrival time. Moreover t_a denotes arrival time and tt is the travel time. Operator can enter some distributions for α (VOT), β and γ (delay penalties) and t^{*}. These behavioral parameters are the only information that are necessary in addition to data required by the classical static traffic assignment models (such as network topology, links characteristics and O-D matrices).

The mode choice is described by a binary logit model. The output of METROPOLIS that we mainly use in this study is the surplus for any traveler category and for any O-D pair. As the departure time choice is modeled by a continuous Logit model, the surplus is given by:

$$S = \mu^T \ln \int exp\left(-C(t)/\mu^T\right) dt,$$

where $C(t)$ is the time dependent travel cost and where μ^T denotes the departure time choice heterogeneity parameter (see also de Palma et al. (1997) and de Palma & Marchal (2002)). For more information about METROPOLIS application and calibration in Ile-de-France, see SustainCity W.P. 6.7.

8 Integrated model of METROPOLIS and UrbanSim

8.1 Integration of METROPOLIS and UrbanSim

An interface is developed to manage the simulation process and data exchange between METROPOLIS and UrbanSim. The interface is described in the Chapter 3.3. The interface computes and updates the O-D matrix based on the simulated population and employment data.

8.2 Runs and Calibration of Integrated Model

The urban model consists of several modules which represent several urban subsystems. As we presented before, these modules are estimated using several types of data. Since that the different sources do not necessary span on the same period (We had to find sometime compromises). Each model has a margin of error which is more or less important and depends among others on the quality of available data. The difference between simulated values and observed ones is greater when the simulated period is longer and when the calibration period is shorter. Moreover the simulation represents the observed phenomenon on the basis of the explanatory variables. However, the future is not always simple extrapolation of the past and we have decided to restrict the number of explanatory variables in order to improve the robustness of our prediction. This partially explains the difference between the simulation results and the observation for periods beyond the estimation period.

We use a 2-step procedure to calibrate UrbanSim. In the first step we use estimated coefficients to calibrate the simulation results as a matter of fact the estimations are based on average values of the coefficients and some sensitivity analysis is required to analyze the results when value of coefficient can vary within the confidence interval of the estimation. The idea of calibration is to determine the value of the coefficient within the confidence interval in order to reproduce in the best possible way the observed data. The second step is concerned with the period beyond the estimation period. In this case, we adjust various constants specific to the geographical units and to the different groups of agents in order to get simulation results coherent with actual data.

In the case of Ile-de-France, the first step corresponds to the period 1990 - 1999 and the second step corresponds to the period 1999 to 2006. We mainly fit the residential location model as well as the establishment location choice model. A constant was introduced for each *CDT* (10 zones), for each type of Household (24 types), and for each business activity (10 sectors). We have obtained a mean relative weighted error equal to 0.07% for the 10 *CDT* and the R^2 is larger than 99% for the employment (simulated versus observed). These are very encouraging results. We believe that the quality of the calibration is good, and then it can be used to study different scenarios for the Grand Paris. This will be discussed below.

9 Policy Scenarios and Results: The Grand Paris Project

Several policy scenarios have been considered to be simulated by the integrated model.Among them, we describe the Grand Paris project that has several particularities that make it the most interesting to present in this chapter.

9.1 The Grand Paris Project

In the long run, the Grand Paris project aims at transforming Paris area in its role as a large world metropolis. The realization of the Grand Paris Project will rely on large economic centers (named *CDTs* for *Contrats de Développement Territorial*) around Paris historical city, connected through a new fast public transportation network (named *RTPGP* for *Réseau de Transport Public du Grand Paris*), and governed by an institutional structure, the *Société du Grand Paris*.

The radial structure of the current transportation network in Paris area favors trips inside and towards Paris, but it is ill-adapted for transversal trips from the suburbs to the suburbs. Currently, most of the trips between two cities located in the suburbs must transit through Paris, which induces an inefficient congestion in the transportation network.

The project analyzed here is made of 3 automatic Metro lines representing a total of nearly 160 km, and 95 stops in 84 stations, for a budget of nearly 30 billion euro. These lines complete and improve the structure and efficiency of the existing network and go through 117 *communes*. The objective is to avoid useless trips through Paris, attract a large enough fraction of suburbs-to-suburbs trips, and thus release the pressure currently affecting the central Paris network.

Two main phases are considered for the realization of the Grand Paris Project, in 2025 and in 2035. The main characteristics of the Grand Paris Metro lines are presented in the Table11.

9.2 Expected impact on urban development and envisaged scenarios

The Grand Paris Project increases by more than 50 percent the total length of Metro lines, and by 20 percent the number of Metro stops in Paris area. The transport supply improvement induced by this project will dramatically improve accessibility across the region, and will considerably improve regional attractiveness and competitiveness. Measuring the aggregate effect of the project at the regional level would

Line	Peak hour trains interval	Line Capacity	Commercial speed
Blue	85 s	38,000	45 km/hr
Red	120 s	32,000	60 km/hr
Green	150 s	6,000	65 km/hr

Table 11: Main characteristics of the Grand Paris automatic Metro lines (Source: SGP 2011)

be out of the scope of this study, and it would require different tools adapted for macroeconomic and interregional analysis. Such macroeconomic studies were financed by the *Société du Grand Paris*, and their results are used as inputs here.

Three scenarios were established concerning population and employment:

- *Reference scenario*. This scenario corresponds to the regional evolution without the Grand Paris Project. It relies on the projections used in the draft of *SDRIF*[9] prepared by the *Institut d'Aménagement et d'Urbanisme de la région Île-de-France (IAU-IdF)*. This scenario predicts 1.348 million inhabitants more and 0.683 million jobs more by 2030.

- *Low scenario* about regional attractiveness. This scenario corresponds to the lower limit of the projections of the effect of the the Grand Paris Project on regional attractiveness. This scenario predicts 1.406 million inhabitants more and 0.8 million jobs more by 2030.

- *High scenario* about regional attractiveness. This scenario corresponds to the upper limit of the projections of the effect of the the Grand Paris Project on regional attractiveness. This scenario predicts 1.5 million inhabitants more and 1 million jobs more by 2030.

Since the phases of the Grand Paris project are scheduled in 2025 and 2035, we translated the above values into projections for 2025 and 2035, and we used a piecewise-linear function for intermediate years. Table 12 presents rounded and exact values for the three scenarios.

The main objective of the study was to evaluate the effects of the Grand Paris Project until year 2035. Additional projections were run until 2050 assuming constant aggregate population and employment, in order to evaluate the long run evolution of the distribution of population and employment within the region after the network and total population and employment remain steady. Such very-long-run projections are less precise because the duration of the projection is very large.

In addition, the Grand Paris Project should improve the local economic development of of some fractions of the region, such as the *départements* Seine-Saint-Denis

[9] *Schema Directeur de la Région Ile-de-France* is the regional master plan for Ile-de-France (**?**).

Scenario	Values	Population Increase			Employment Increase		
		2025	2030	2035	2025	2030	2035
Ref.	Round.	1,100,000	1,300,000	1,600,000	500,000	700,000	800,000
	Exact	1,078,966	1,348,706	1,618,446	546,825	683,535	820,245
Low	Round.	1,100,000	1,400,000	1,700,000	600,000	800,000	1,000,000
	Exact	1,125,070	1,406,322	1,687,573	639,996	799,996	959,995
High	Round.	1,200,000	1,500,000	1,800,000	800,000	1,000,000	1,200,000
	Exact	1,199,997	1,500,006	1,800,014	800,004	1,000,006	1,200,008

Table 12: Input data for the 3 scenarios (Source: *SGP, DRIEA*)

or the *Plateau de Saclay*. The local development in those sub-regions is currently restricted by the lack of an efficient transportation network.

9.3 Data preparation for the 3 scenarios

Yearly control totals for the 3 scenarios, over the period 1999-2050, were required to run UrbanSimE. To perform this exercise, we used and extrapolated the input data provided by the *Société du Grand Paris* for the 3 scenarios in 2005, 2025, 2030, 2035 and 2050, as well as census data and other data provided by *INSEE*. The yearly control totals and intermediary variables used to compute them are:

- Total population and number of households, distributed over the 24 types chosen and total employment, by activity sector;
- Transportation supply and service levels:
  - Matrices of flow, distances and travel times, by mode, at each period, under each scenario;
  - Graphs of road and public transport networks, metro and rail stations and stops, at each period, under each scenario.

9.3.1 Total population

We extrapolated the total population from the input data described in Table 13. This input data comes from Census and *INSEE* for the past, and comes through *SGP* for the projections.

For the yearly extrapolation of total population, we assumed a constant yearly growth over each period between two lines in Table 13. This yearly growth rate is specific to each scenario after 2008. Each growth rate was computed to fit the figures of Table 13, and is reported in Table 14. The resulting population control totals are represented on Figure 6.

Year	Population			Source
	Reference	Low	High	
1999	10,952,011	10,952,011	10,952,011	Census
2005	11,433,302	11,433,302	11,433,302	*INSEE*
2006	11,532,409	11,532,409	11,532,409	Census
2008	11,659,620	11,659,620	11,659,620	Census
2025	12,512,268	12,558,372	12,633,299	*IAU, SGP*
2030	12,782,008	12,839,624	12,933,308	*IAU, SGP*
2035	13,051,748	13,120,875	13,233,316	*IAU, SGP*
2050	13,051,748	13,120,875	13,233,316	*IAU, SGP*

Table 13: Input data used to extrapolate total population (Source: Census, *INSEE, IAU-IdF, SGP*)

Period	Population growth rate		
	Reference	Low	High
1999 – 2005	0.72%	0.72%	0.72%
2005 – 2006	0.87%	0.87%	0.87%
2006 – 2008	0.55%	0.55%	0.55%
2008 – 2025	0.45%	0.47%	0.51%
2025 – 2030	0.43%	0.44%	0.47%
2030 – 2035	0.42%	0.43%	0.46%

Table 14: Population growth rate, by period and scenario (Source: authors' computations)

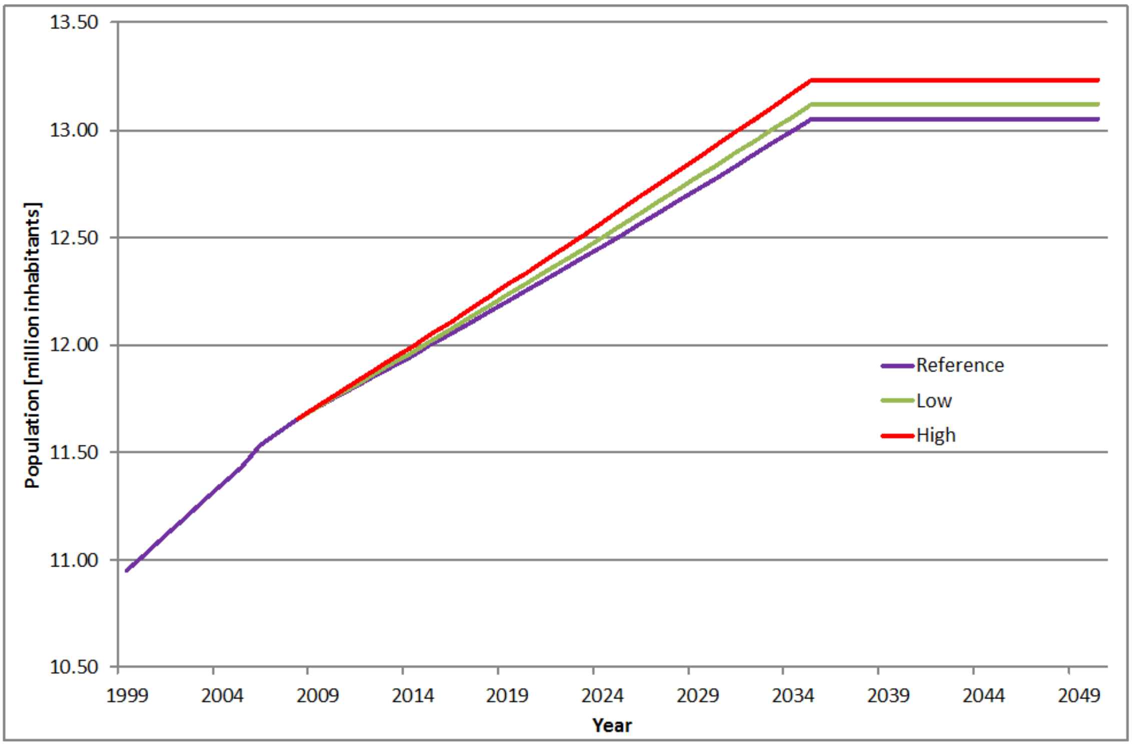


Figure 6: Control totals for population, 1999-2050 (Source: Census, authors' computations)

The decrease in growth rate between the two periods 1999-2006 and 2006-2008, both concerning census data, probably result more from changes in census data collection method (from exhaustive census each 10 year to partial yearly census) than from real changes in population growth rates. Such inconsistencies in our input data do not affect our simulations.

9.3.2 Number and size of households

We translated the yearly population control totals into number of households, by size (and type). In order to do this, we took into account and extrapolated to the period 2008-2035 the decreasing trend in household size observed over the period 1999-2008. The evolution of the distribution of households by size is assumed the same for the 3 scenarios. The resulting average household size is represented on Figure 7. The yearly number of households is obtained as the ratio between yearly population and yearly average size. The evolution of the number of households by type is specific to each scenario, like the evolution of total population.

We then computed the yearly distribution of households of by size, extrapolating the trend observed in 1999-2006. This distribution is represented in Table 15 for the

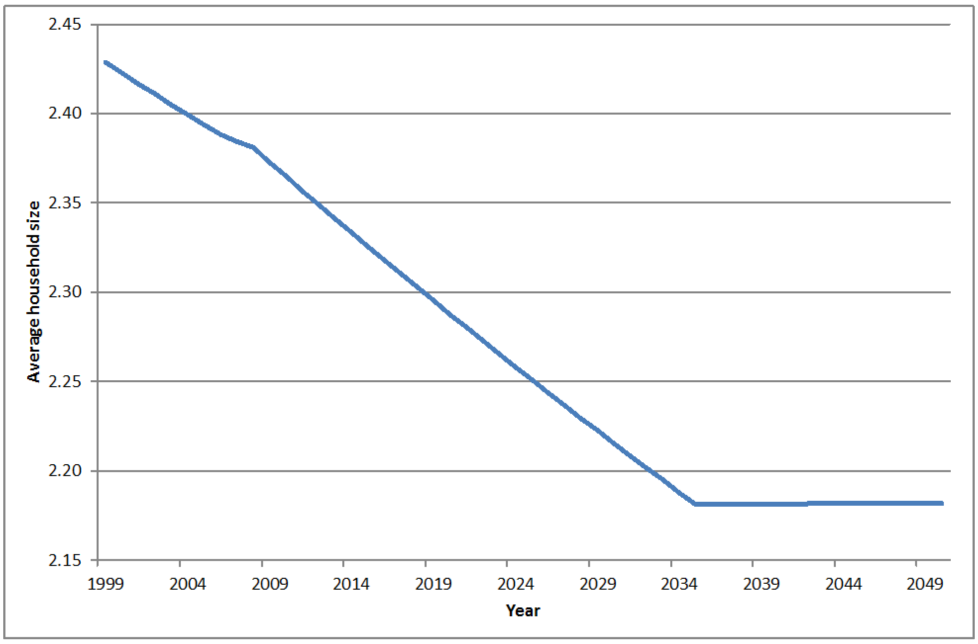


Figure 7: Average household size over the period 1999-2050 (Source: Census, authors' computations)

key years, and in Figure 9 for the yearly evolution.

Finally, households are distributed by dwelling type and tenure status, conditional on their size. This defines 24 types (6 size categories * 2 dwelling types * 2 tenure status categories). To do this, we assumed that the distribution of household dwelling/tenure types conditional on household size is constant over time, and equal to the distribution observed in 2008.

Year	Distribution of household size						Average size	
	1	2	3	4	5	6+	6+	Total
1999	0.346	0.279	0.157	0.136	0.054	0.028	8.539	2.429
2006	0.354	0.287	0.149	0.130	0.053	0.026	8.601	2.388
2025	0.381	0.304	0.133	0.114	0.049	0.019	8.614	2.251
2035	0.397	0.311	0.124	0.106	0.046	0.015	8.644	2.181
2050	0.397	0.311	0.124	0.106	0.046	0.015	8.644	2.181

Table 15: Distribution of household size and average size, at key periods (Source: Census, authors' computations)

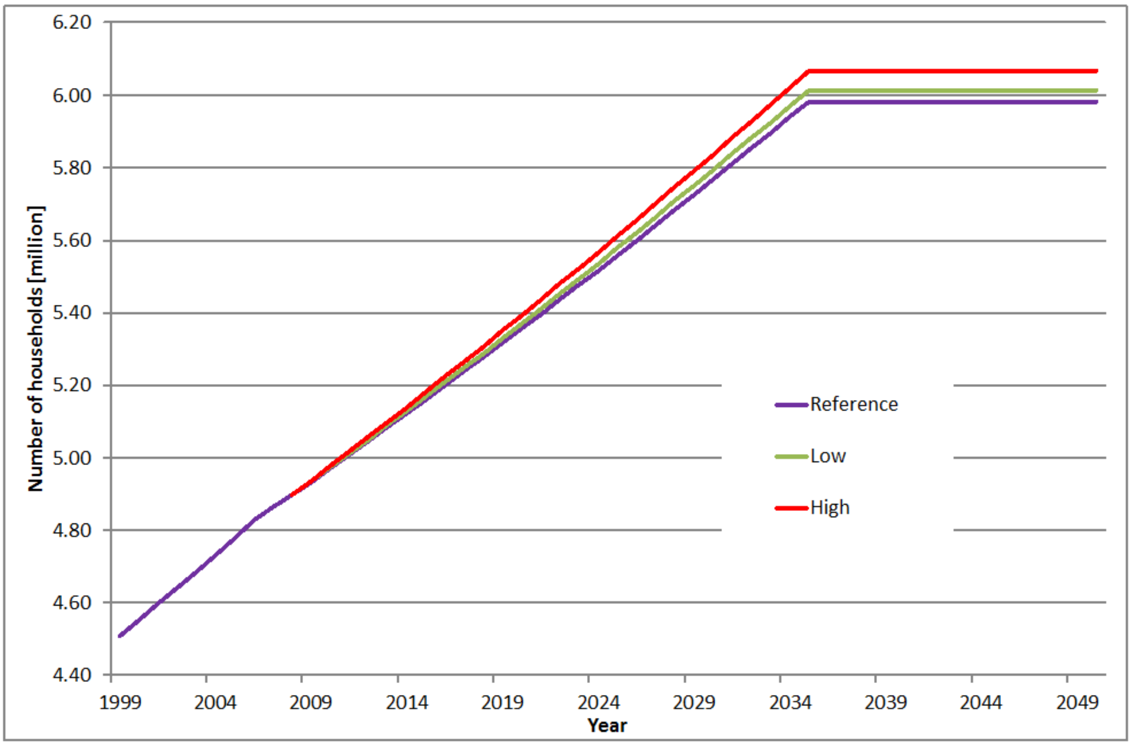


Figure 8: Total number of households for period 1999-2050 (Source: Census, authors' computations)

9.3.3 Total employment

A methodology similar to the one used for population was applied to build employment control totals:

- First step: Build control totals for total employment over the period 1999-2050;
- Second step: Distribute the yearly control totals for total employment over the 11 activity sectors considered.

The extrapolation of total employment control totals relies either on input values obtained through *SGP* (for projections), or from census data (for past data). Table 16 sums up the information available at key dates and the assumptions used to extrapolate this information to yearly control totals, and in Figure 10 represents the yearly evolution of total employment control totals over the period 1999-2050.

Although we are aware that the distribution of employment across activity sectors evolves across time, we have not access to relevant data to measure this evolution in the calibration period 1999-2008, and extrapolate it in the projection period.

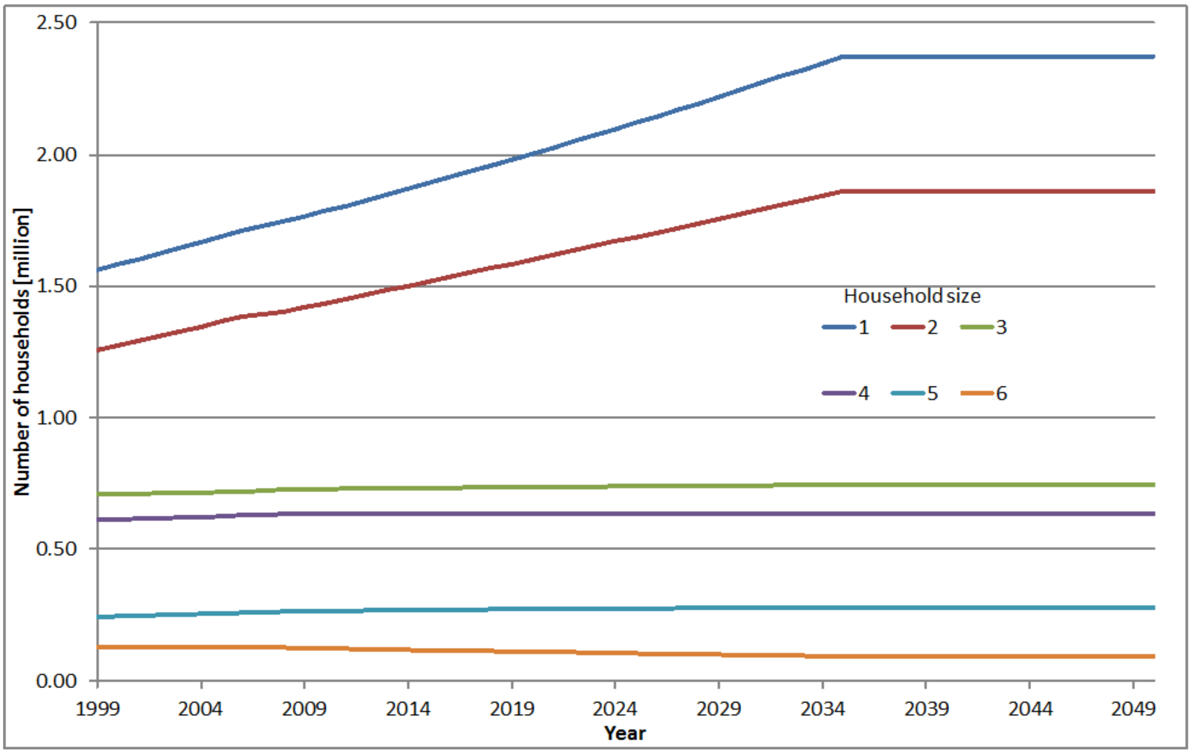


Figure 9: Number of households by size category for period 1999-2050 (Source: Census, authors' computations)

This is partly due to a significant change in the typology of jobs used to provide employment data, which makes is difficult to follow sectoral distribution of employment over time. We therefore have to rely on simplifying hypotheses, and assumed that the sectoral distribution of employment aggregated at the regional level is constant over time.

9.3.4 Transportation supply

The improvement of the public transportation network will directly reduce public transportation travel times on the lines concerned. This will have different indirect effects: Some individuals will change their mode from car to public transit as a response to improvements in public transportation infrastructure, which will reduce congestion on the roads; other individuals will possibly change their mode from public transit to car as a response to congestion reduction (this second effect should be small compared to the first one); some individuals will change residential and/or job location as a response to accessibility changes resulting directly from the infrastructure and indirectly from changes in mode choice. This will induce a significant change in the O-D matrix of flows and, consequently, in the O-D matrix of travel

Year	Number of jobs			Source
	Reference	Low	High	
1999	5,041,995	5,041,995	5,041,995	Census
...	Constant growth rate			
2005	5,360,447	5,360,447	5,360,447	Census
...	Constant growth rate			
2025	5,907,272	6,000,443	6,160,451	*SGP*, input data
...	Linear evolution			
2035	6,180,692	6,320,442	6,560,455	*SGP*, input data
...	Constant			
2050	6,180,692	6,320,442	6,560,455	*SGP*, input data

Table 16: Input values and assumptions for extrapolating total employment (Source: Census, *DRIEA, SGP*)

times. METROPOLIS computes the equilibrium resulting from such interactions, and produces new O-D matrices of travel times by private car.

DRIEA provided O-D matrices of travel times and flows computed from the 4-step transportation model *MODUS* at key dates, for each scenario, as indicated in Table 17, which also presents the assumptions used to extrapolate these matrices between the key dates. These matrices were obtained under assumptions about household and job location which may a priori not be consistent with our own results. However, the differences in the distribution of household and job locations between the data provided by *DRIEA* and the output of METROPOLIS are not very large, so such potential inconsistencies are limited.

DRIEA provided O-D matrices at the *MODUS zones* level. There are 1289 *MODUS zones*, and the inputs and outputs of METROPOLIS also defined at the *MODUS zones* level. In the dense part of Paris area, *MODUS zones* mainly correspond to IRIS zones or groups of several IRIS included in the same *communes*, whereas *MODUS zones* mainly correspond to *communes* or groups of *communes* in the less dense part of Paris area. We built a correspondence between the 1289 *MODUS zones* and the 1300 *communes* of Paris area. We used this correspondence to build O-D matrices of flows and travel times at the *commune* level.

Matrices of flows, travel times and distances correspond to weighted average values of the same variables, averaged over all the itineraries between the relevant zones. The weights correspond to the share of the itineraries in the relevant O-D pair.

In combination with the population and employment scenarios defined in Section 9.2, we consider two different transport scenarios: The *Reference* scenario does not benefit from the Grand Paris Project infrastructure, whereas the *Grand Paris* (or *GP*) scenario does. In the results section, the *reference* scenario will correspond

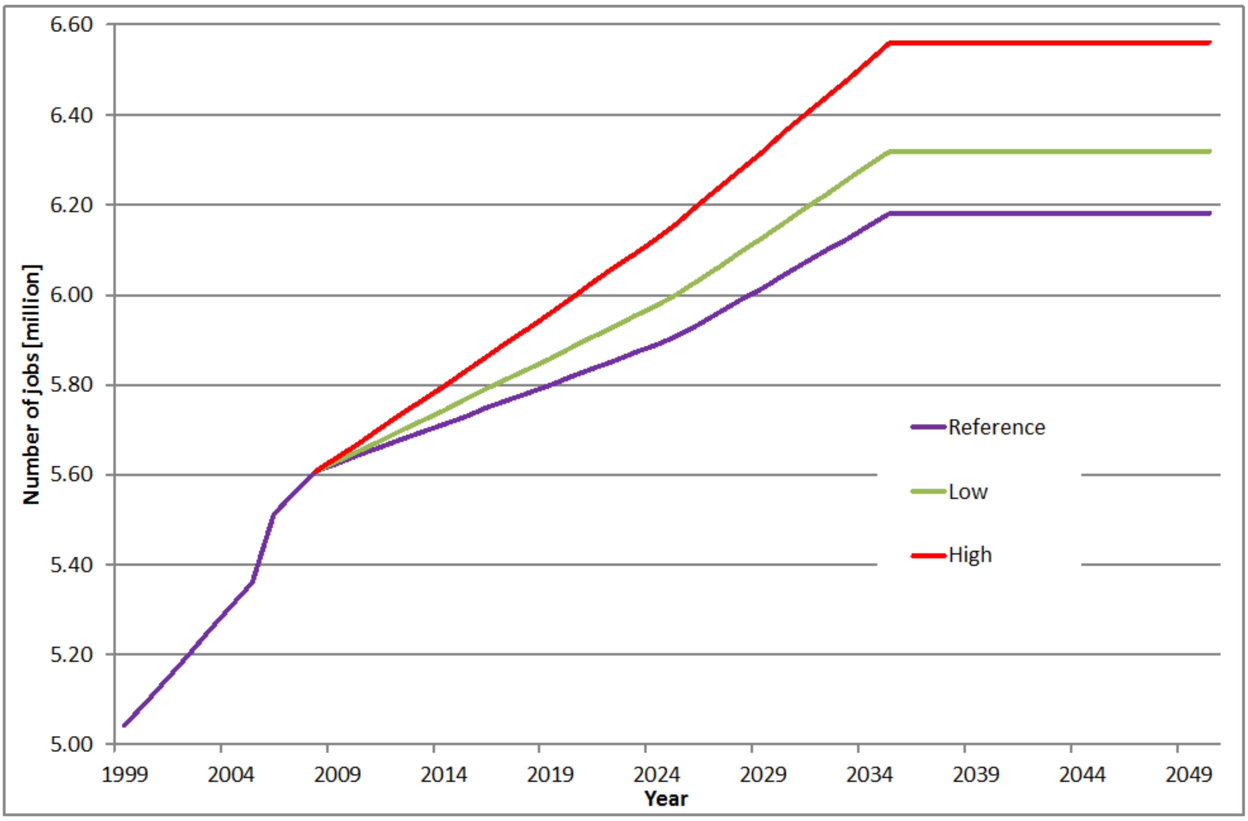


Figure 10: Total number of jobs for period 1999-2050 (Source: Census, authors' computations)

to the reference both for population, employment and transportation; the *Ref.-GP* scenario will correspond to the *reference* scenario for population and employment, but to the *Grand Paris* scenario for transportation, whereas only the *Grand Paris* transportation scenario will be considered in the (population and employment) *Low* and *High* scenarios.

For the simulations with METROPOLIS, the population and employment data produced by UrbanSim and introduced in METROPOLIS via interface. The road network characteristics are updated according to the put in operation planning of different projects.

9.4 Results

9.4.1 Evolution of the population

The aggregate population levels were given in the description of the scenarios in Section 9.2. Table 18 presents the population increase between 2005 and 2035 in the different scenarios, in Paris, in each CDT, and in the rest of the region. Figure

<table>
<tr><th>Year</th><th>Ref: Public</th><th>Ref: Private</th><th>GP: Public</th><th>GP: Private</th></tr>
<tr><td>1999</td><td rowspan="3">MODUS 2004</td><td rowspan="3">MODUS 2003</td><td rowspan="3">MODUS 2004</td><td rowspan="3">MODUS 2003</td></tr>
<tr><td>2003</td></tr>
<tr><td>2004</td></tr>
<tr><td>2005</td><td rowspan="2">MODUS 2005</td><td rowspan="2">MODUS 2009</td><td rowspan="2">MODUS 2005</td><td rowspan="2">MODUS 2009</td></tr>
<tr><td>...</td></tr>
<tr><td>2020</td><td rowspan="3">IAU-SDRIF 2025</td><td rowspan="3">IAU-SDRIF 2025</td><td>GPX-REF 2025</td><td>GPX-REF 2025</td></tr>
<tr><td>...</td><td colspan="2">Evolution at a constant rate</td></tr>
<tr><td>2025</td><td>GPA 2025</td><td>GPA 2025</td></tr>
<tr><td>...</td><td colspan="4">Evolution at a constant rate</td></tr>
<tr><td>2035</td><td rowspan="3">IAU-SDRIF2035</td><td rowspan="3">IAU-SDRIF2035</td><td rowspan="3">GPX 2035</td><td rowspan="3">GPX 2035</td></tr>
<tr><td>...</td></tr>
<tr><td>2050</td></tr>
</table>

Table 17: Data sources at key dates and assumptions for extrapolating O-D matrices

11 represents the evolution across time of the fraction of the cumulated increase in population (compared to 2005) which goes to the historical Paris, to the secondary centers (the *CDTs*, which are mainly located in the close suburbs and in some relatively dense parts of the far away suburbs), and to the core of the region, which is made of Paris plus CDTs.

The negative figures for Paris until 2030 means that, whatever the scenario, population inside Paris will decrease, whereas the regional population will increase. Such a decrease was already observed in 2008 according to census data. In 2008, the decrease in the share of Paris in regional population growth was exactly compensated by an increase in the CDTs, so that the population in the core of Paris area is the same in 2008 as it was in 2005. The population decrease in Paris results from the conjunction of two phenomena: On the one side, strong capacity constraints inside Paris making it very difficult to build new dwellings; on the other side, the decreasing trend in household size implies a decrease in population when the number of households remains constant. Our simulation results suggest that the share of Paris in total population will remain stable until 2020, and will then increase until 2035, with small differences between scenarios.

Our results also suggest a decrease in the fraction of the additional population which goes to CDTs from 2008 to 2015 in all the scenarios. This reflects and extrapolates the current trend of the population to move from the core of the city to the far away suburbs. This is because the CDTs currently suffer both from high real estate prices and relatively poor accessibility. From 2015 to about 2020, we observe a very small increase in the share of CDTs in the *High* and *Low* scenarios, and a stabilization of this share in the in the *Reference* scenario, either with or without the *Grand Paris* infrastructure. In the *High* and *Low* scenarios, population growth

Horizon	2035				Input data, *High*
Scenario	*Reference*	*Ref.-GP*	*Low*	*High*	
Zone					
Region	1,618,446	1,618,446	1,687,573	1,800,014	1,800,003
Paris +CDTs	**623,846**	**861,361**	**954,312**	**1,009,113**	**1,186,986**
Rest of the region	994,600	757,085	733,261	790,901	613,017
Paris	**93,229**	**89,555**	**89,586**	**90,661**	**141,042**
Aulnay-Montfermeil	23,780	37,929	47,548	58,101	96,084
Biotechnologies S.-A.	98,645	154,814	175,079	184,985	206,145
Confluence	65,128	52,653	51,075	54,058	49,525
Descartes	49,337	67,795	77,875	86,034	104,132
La Défense	56,339	71,423	78,514	83,531	98,487
Le Bourget	18,778	71,964	87,449	89,014	102,000
Pleyel	72,249	112,897	123,706	128,930	102,666
Roissy-pôle	28,965	29,409	30,560	32,291	58,033
Saclay	100,040	156,522	175,761	182,764	223,313
Val de France - Gonesse	17,356	16,400	17,159	18,744	5,559

Table 18: The evolution of population between 2005 and 2035 (Source: simulation results)

in the CDTs will anticipate the Grand Paris transportation infrastructure and start locating in the places which will become more accessible a few years before accessibility actually improves. This can be explained by urban policies aimed at favoring housing construction, which will start several years before the transportation infrastructure actually works. This phenomenon starts later (only in 2019, which is only 1 or 2 years anticipation) in the reference scenario for population and employment, with the Grand Paris infrastructure, because urban policies are less needed and will therefore be less active in this scenario.

In the long run, the CDTs attract 56% of the increase in population in the Low and High scenarios, and 52% in the reference scenario with Grand Paris infrastructure, whereas this fraction is only 35% in the reference scenario. The improvement of accessibility and urban policies favoring the construction of dwellings in the CDTs attract the majority of the population in the CDTs.

This fraction goes up to 61% for the core of the region in the Low and high scenarios, 58% in the reference scenario with Grand Paris infrastructure, and 40% in the reference scenario.

In the intermediary period 2020-2035, the rhythm of population growth in the CDT relatively to the rest of the region significantly varies across scenarios. In the reference scenario with Grand Paris infrastructure, the relative growth started later than in the Low and High scenarios (2019 instead of 2015) because total population growth is less important, so there is less need to anticipate the construction of the infrastructure. Relative population growth rate in the reference scenario with Grand Paris infrastructure then rapidly catches up the gap compared to the Low and High scenarios between 2019 and 2023. The infrastructure is anticipated later when total population is smaller, but the difference disappears just before the first phase of the project ends. The relative evolution of population growth is then the same in the three scenarios with Grand Paris infrastructure until 2026, which means that it does not depend on the size of the aggregate population in the region during this period. Finally, the relative share of the population in the CDTs grows faster in the period between the two phases of the construction of the infrastructure (2026-2035) in the Low and High scenarios than in the reference scenario with Grand Paris infrastructure. This may indicate that the second phase of the project is needed to attract the additional aggregate population assumed to come to the region in the Low and High scenarios compared to the reference scenario. Without the second phase of the project, this additional population would induce too much congestion in the CDTs if it attracted the same share of the growth in Low and High scenarios than in the reference scenario. Stated differently, the first phase of the project is sufficient to attract population in the CDTs in the reference population scenario, whereas the second phase is needed in the Low and High scenarios.

9.4.2 Evolution of Employment

The aggregate employment levels were given in the description of the scenarios in Section 9.2. Table 19 presents the employment increase between 2005 and 2035 in the different scenarios, in Paris, in each CDT, and in the rest of the region. Figure 12 represents the evolution across time of the fraction of the cumulated increase in employment (compared to 2005) which goes to the historical Paris, to the secondary centers (the *CDTs*), and to the core of the region.

The evolution of employment in Paris is subject to capacity constraints similar to those applying to the population, but the trends in the needs are totally different. Whereas each individual tends to require more and more dwelling capacity (because of the decreasing trend in household size), each job requires less floor space because of the decreasing trend in surface per worker. This decreasing trend allows the number of jobs to increase inside Paris from the beginning of the simulation period although the floor space devoted to employment does not increase before year 2020, except for some projects already planned, and introduced exogenously in the system. Capacity constraints and the decreasing trend in the surface per worker are the main forces driving the evolution of the number of jobs, which leaves no room for economic considerations. Strong capacity constraints in Paris imply that the jobs which locate inside Paris are those that can do so, not all that want to do so, which are more numerous in all the scenarios with the Grand Paris infrastructure.

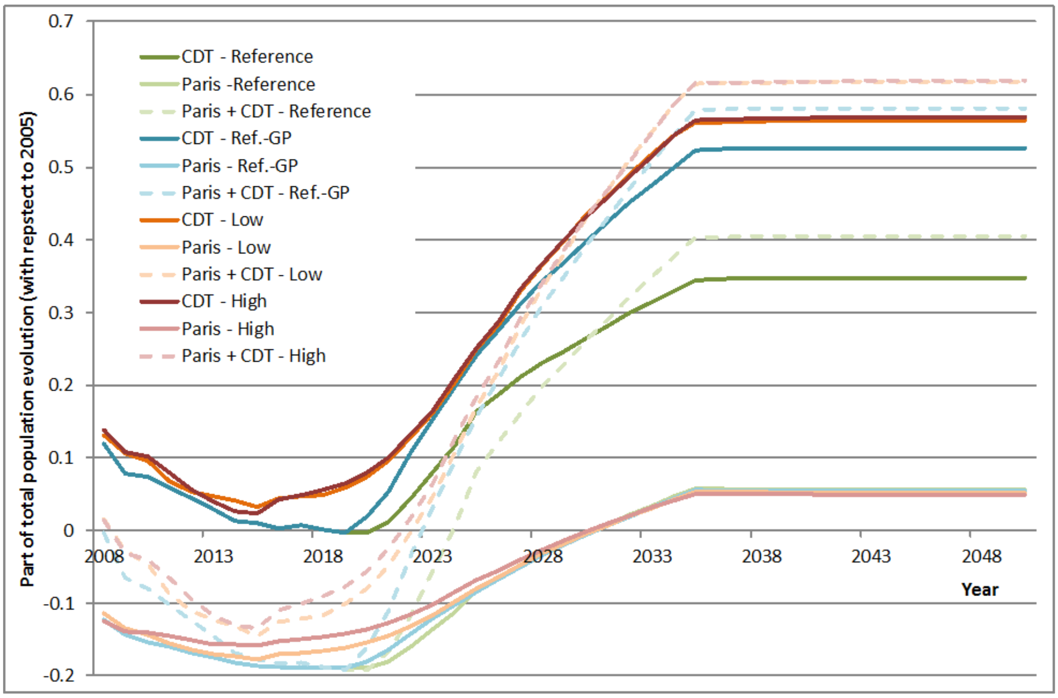


Figure 11: Evolution of the population for period 1999 to 2050 (Source: simulation results)

When comparing High, Low and reference scenarios with Grand Paris project, the share of Paris in total employment decreases when the size of total regional employment increases, but the number of jobs located inside Paris in these three scenarios does not differ significantly. The release in capacity constraints inside Paris allowed by the decrease in floor space needed per job explains that the share of the CDTs slightly decreases between 2035 and 2050 to the benefit of Paris, the share of the core remaining constant over this period.

The capacity constraints mentioned above were short run constraints, i.e. without changes in land use type. In the reference scenario without the Grand Paris project, the demand for jobs to locate inside Paris is significantly larger than in the reference scenario with the Grand Paris project. This is because, when the infrastructure is built but the aggregate regional employment is the one of the reference scenario, improvements of accessibility will attract many jobs to the CDTs and decrease the demand for jobs to locate inside Paris. This excess demand for jobs inside Paris in the reference scenario will be so large that it will move upwards the short run capacity constraints: The economic pressure will be so strong in the reference scenario that it will force some changes in land use type and construction of additional office buildings. As a consequence, the number of jobs located in Paris, and the relative share of Paris in the increase of regional employment are larger in the

Horizon	2035				Input data, *High*
Scenario	*Reference*	*Ref.-GP*	*Low*	*High*	
Zone					
Region	820,245	820,245	959,995	1,200,008	1,199,999
Paris +CDT	**495,723**	**650,979**	**776,876**	**927,578**	**1,010,657**
Rest of the region	324,522	169,266	183,119	272,430	189,342
Paris	**112,138**	**91,678**	**90,575**	**94,430**	**74,339**
Aulnay-Montfermeil	2,381	14,552	19,021	21,260	23,019
Biotechnologies S.-A.	75,856	69,868	90,885	128,976	144,836
Confluence	28,296	26,446	27,116	28,931	17,885
Descartes	39,719	60,782	69,696	79,001	84,246
La Défense	96,097	128,586	139,625	143,918	146,794
Le Bourget	15,092	33,629	42,416	49,640	60,000
Pleyel	55,980	75,183	85,043	93,163	115,489
Roissy-pôle	29,014	53,369	84,130	126,819	170,841
Saclay	35,187	94,820	121,294	144,125	159,019
Val de France - Gonesse	5,963	2,064	7,075	17,315	14,189

Table 19: Evolution of the number of jobs between 2005 and 2035 (Source: simulation results)

reference scenario without Grand Paris infrastructure than with this infrastructure.

Compared to the population, the final share of employment which goes to the CDTs and to the core of Paris area, and the pace at which these shares evolve across time slightly differ between Low and High scenarios. The total number of additional people working in CDTs or in the core are larger in the High scenario than in the Low scenario, but the share of new jobs going to CDTs or to the core is smaller (and its evolution is slower between 2022 and 2030) in the High scenario than in the Low scenario. This suggests that more jobs could be attracted by the CDTs and by the core of the region if an even more ambitious infrastructure was built in phase 2. Stated differently, the improvement in accessibility is not enough to attract the full potential of jobs which could be attracted by CDTs and the core of Paris area in the High scenario.

Nevertheless, more than 70% of the additional jobs will be attracted by the CDTs (more than 80% by the core of Paris area) if the Grand Paris infrastructure is built, whereas CDTs would attract only 47% (and the core 60%) of the new jobs in the reference scenario without Grand Paris infrastructure.

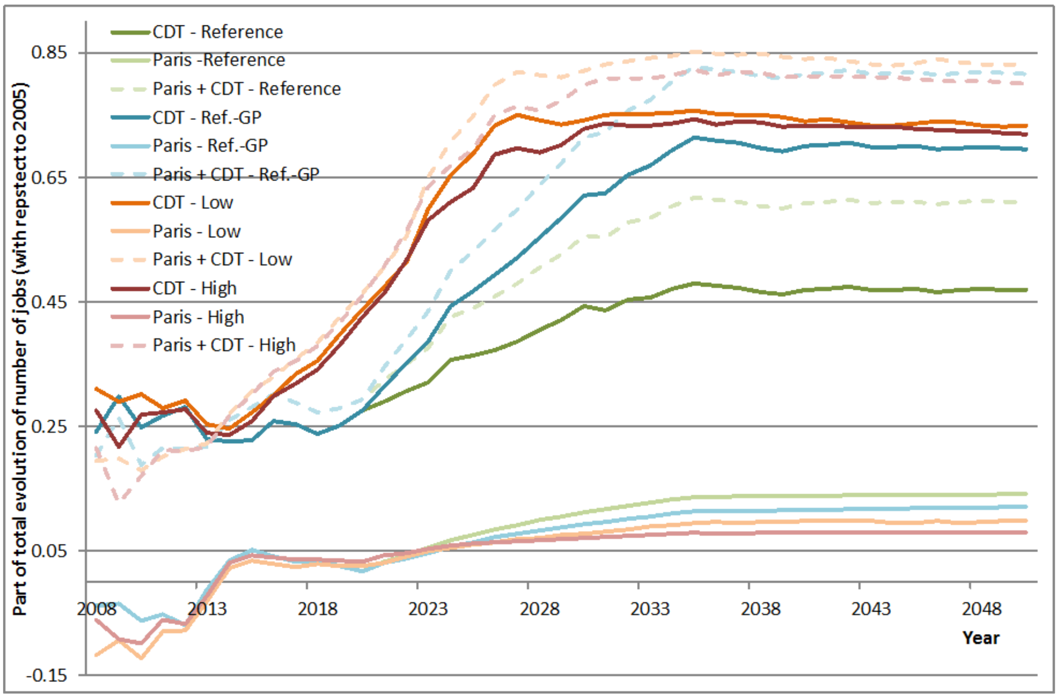


Figure 12: Evolution of the number of jobs for period 1999-2050 (Source: simulation results)

9.4.3 Evolution of inequalities within the region

Based on UrbanSimE simulation results, we analyzed the evolution of of inequalities, expressed in terms of income per capita and in demographic terms. Since the household income is not observed directly in census data, we predicted it, using the French Family Budget Survey (**?**). Income is predicted as a function of household characteristics such as size (by age group), number of working members, age, activity, diploma, and nationality of household head. In order to take into account economies of scale, the household income is divided by the square root of the number of household members, which is an equivalence scale commonly used. Without any equivalence scale, Paris would appear poor compared to the rest of the region just because households are smaller in Paris than in the rest of the region, and more generally, the apparent ranking of *commune* or *départements* by income level would be inconsistent. This problem is solved by our equivalence scale, and the ranking of *commune* or *départements* by income level is totally consistent.

An equivalence scale is intended to reflect the way the needs vary with household size. With this equivalence scale, the economic well-being of a household with one member and 2,000 euro per month is the same as the economic well-being of a household with 4 members and 4,000 euro per month. The distribution of income per capita is assumed constant over the period, at the region level. As a

Variable	**Percentage Rich hhs**			**Percentage Poor hhs**			**Poor-to-Rich ratio**		
Year	**2035**			**2035**			**2035**		
Scenario	**Ref.**	**Low**	**High**	**Ref.**	**Low**	**High**	**Ref.**	**Low**	**High**
Other *communes*	34.7	34.7	34.5	28.7	28.6	28.6	0.83	0.82	0.83
Paris	23.7	23.8	23.7	41.5	41.4	41.4	1.75	1.74	1.75
Aulnay-Montfermeil	19.4	19.0	19.5	40.6	41.1	41.4	2.09	2.17	2.13
Biotechnologies S.-A.	28.6	30.5	30.6	33.1	31.8	31.7	1.16	1.04	1.04
Confluence	36.5	35.9	36.8	27.1	27.2	26.6	0.74	0.76	0.72
Descartes	33.0	33.2	34.2	27.8	27.6	27.3	0.84	0.83	0.80
La Défense	37.0	37.6	37.5	26.2	25.9	26.0	0.71	0.69	0.69
Le Bourget	12.2	12.4	14.4	55.6	54.5	53.4	4.56	4.40	3.72
Pleyel	9.6	9.8	9.9	60.3	60.0	60.1	6.31	6.12	6.05
Roissy-pôle	26.9	28.8	29.5	32.6	30.7	31.1	1.21	1.07	1.05
Saclay	42.7	44.1	43.5	22.0	21.3	21.6	0.52	0.48	0.50
Val de France - Gonesse	15.9	15.6	16.0	45.5	45.9	45.9	2.86	2.95	2.87
Region	30.7	30.7	30.7	33.1	33.1	33.1	1.08	1.08	1.08

Table 20: Distribution of income and Poor-to-Rich ratio (Source: simulation results)

consequence, the income considered here has to be interpreted as a relative income (relative to the average regional level). This normalization has no consequences on our simulation results. We simply assume that the distribution of the population over the region would be the same if the income of each household increased by 10%. This assumption does not seem really restrictive.

Income per capita is divided in 3 classes (Rich, middle income and poor), built so that each of them represents about a third of the sample (actually, there are 30.7% Rich households, 33.1% Poor households and 36.2% Middle Income households in the region). There is a strong tendency to the increase of intra-regional inequalities over time, whatever the scenario.

Paris tends to become poorer as the richest households will tend to move to the suburbs, whereas the poorest ones will tend to stay inside Paris. As a result, in all the scenarios, there will be more than 41% poor households and less than 24% rich households inside Paris in 2035, whereas there were only 36.7% poor households and as much as 31.5% rich households inside Paris in 1999. The change in the income distribution is less severe in the other zones, although it systematically goes in the direction of an increase of inequalities over time. Outside Paris, the richest zones tend to become even richer, and the poorest zones tend to become poorer.

We computed the Poor-to-Rich ratio as an indicator of inequalities, and we represent this ratio in Table 12 for each zone in each scenario.

Overall, the Grand Paris project marginally reduces intra-regional inequalities in income compared to the Reference scenario. Income distribution and differences in poverty are not significantly affected by it, although there are large differences between zones in all the scenarios. The poorest zones such as Pleyel or Le Bourget remain the poorest ones in the region, but the fraction of Poor households is marginally reduced in those CDTs.

10 Conclusions

The Paris case study in SustainCity has profited from several years of research in UrbanSim application in Ile-de-France. The data have been gathered over several years and they were updated and enriched with new sources within SustainCity framework. As urban modeling and in particular microsimulation urban models need a huge volume of data that often are not completely available, the model structure and modeling strategy is driven by available data. On the other hand, more data with a better quality become available and impose us to update the models with new possibilities.

The most important task, once data have been collected is the calibration task. The UrbanSim modules have been calibrated separately against observed data in particular the household and individual population data and employment data by activity sector. To present the results and calibrate the model, we have used an aggregated zoning with 50 zones over the region. These zones are designed to be compatible with urban policy zones and also transportation analysis zones. The compatibility with the urban policy zones allowed us to introduce some calibration constants that represent the effects of policies that are not explicitly modeled. The UrbanSimE and integrated model results have been calibrated against the census data for the period 2006 – 2008 for population and employment. We used UrbanSimE in order to evaluate the local effect of the Grand Paris Project on the location of jobs and population within the region, and on real estate prices. UrbanSimE is indeed very efficient in taking into account the heterogeneity of population and firms responses to large infrastructure projects, and in reproducing the dynamics of the evolution induced by such large projects. More generally, microsimulation models like UrbanSimE are well suited for evaluating the intra-regional micro-economic effects of large infrastructure projects, whereas general equilibrium models are well suited for evaluating the inter-regional macro-economic effects. These two types of models were successfully combined for evaluating the effects of the Grand Paris project. The results of calibration procedure presented in Section 8, show that the fit with collected data is very good. The forecasts we have made (and discussed in Section 9) were presented to the *Société du Grand Paris*, who is currently using this integrated model. On line reports provide the results which are often publicly challenged during animated debates. This is, after all, the ultimate scope of large scale models: Challenge intuition and occasionally, present new results and intuition.

In Sustain City, there are two other case studies: Brussels and Zurich. These three case studies are very different since they focus on different cities and are based on different sub-models (residential location models are different for example), market clearing in Paris and Zurich are also different, and the like.

Of course, a model is never finished. One of the reasons, is that any perfect model is likely to become poor after a few years because the behaviors change, and because the technology changes. Stability of behavior is challenged by various economists who advocate the study of endogenous preferences.

In fact, the pure cost-benefit analysis of a project such as the Grand Paris, must be on a minimum of 50 years, or more. We may encounter qualitative changes, which are very difficult to forecast. For example: How long the 2008 crisis will last in France? Another implicit assumption of stationarity made in our approach is related to the technology. For example, the electric car that can change mobility, but we cannot ignore other forms of organization of mobility. We list those that seem most significant: Car-sharing, collective taxis, transport exploiting synergies between individual and collective transport, modular toll, deregulation of taxis, teleworking, teleeducation, etc. Obviously, these management methods will play a major role in mobility and in particular in residential location, in location of business, in fixing the price of land, and the use of space. Researchers still have serious work ahead of them.

References

Anas, A. & Liu, Y. (2007), 'A regional economy, land use and transportation model (relu-tran): Formulation, algorithm design, and testing', *Journal of Regional Science* **47**(3), 415–455.

Anderson, S. P., de Palma, A. & Thisse, J.-F. (1992), *Discrete choice theory of product differentiation*, MIT Press, Cambridge, MA, USA.

Ben-Akiva, M. & Lerman, S. R. (1985), *Discrete choice analysis : theory and application to travel demand*, MIT Press, Cambridge, MA, USA.

Combes, P. P., Lafourcade, M., Thisse, J.-F. & Toutain, J.-C. (2011), 'The rise and fall of spa-tial inequalities in france: A long run perspective', *Explorations in Economic History* (48), 243–271.

de Bok, M. & Sanders, F. (2005), Firm location and the accessibility of locations: Empirical results from the netherlands, *in* 'TRB Annual Meeting', Transportation Research Board.

de Palma, A. & Fontan, C. (2001), 'Choix modal et valeurs du temps en ile-de-france', *Recherche Transports Sécurité,* (71), 24–48.

de Palma, A. & Marchal, F. (2002), 'Real cases applications of the fully dynamic metropolis tool-box: an advocacy for global large-scale mesoscopic transportation systems', *Networks and Spatial Economics* **2**(4).

de Palma, A., Marchal, F. & Nesterov, Y. (1997), 'Metropolis: Modular system for dynamic traffic simulation', *Transportation Research Record: Journal of the Transportation Research Board* **1607**(1), 178–184.

de Palma, A., Motamedi, K., Picard, N. & Waddell, P. (2005), 'Incorporating land use in metropolitan transportation planning', *European Transport* **31**, 67–82.

DREIF (2005), Enquête globale de transport 2001, Technical report, DREIF.

McFadden, D. (1981), Econometric models of probabilistic choice, *in* M. D. Manski, C., ed., 'Structural Analysis of Discrete Data with Econometric Applications', Academic Press.

McFadden, D., Tye, W. & Train, K. (1974), Conditional logit analysis of qualitative choice behavior, *in* 'Frontiers in Economics', Academic Press.

Picard, N. & Antoniou, C. (2011), Econometric guidance, Technical report, SustainCity European Project.

SGP (2011), Schéma d'ensemble, Technical report, Société du Grand Paris.

THEMA-TT-R (2002), QUATUOR: Outils dynamiques de simulation pour la gestion des déplacements en Ile-de-France, Technical report, DRAST/ PREDIT.

Waddell, P. & Ulfarsson, G. F. (2003), Accessibility and agglomeration: Discrete-choice models of employment location by industry sector, *in* 'TRB Annual Meeting', Transportation Research Board.

Acknowledgments

We wish to acknowledge the contribution of several other researchers and students in the implementation of UrbanSimE application in Paris Case study. First of all the support of Paul Waddell and UrbanSim team at University of Washington and University of California, Berkley. Then our colleagues at University of Cergy-Pontoise and *ENS*-Cachan in particular Hakim Ouaras, Navid Khademi, Nicolas Coulombel, Seghir Zerguini, Mohammad Saifuzzaman, Sophie Dantan and Ignacio Inoa. The policy analysis would not have been possible without active support of the *Société de Grand Paris* and in particular of Jean-Claude Prager. Public agencies: *DRIEA* and *IAU-IdF* have provided essential data.